
Statistische Beurteilung der Güte von analytischen Ergebnissen - Untersuchungen zur Präzision in der HPLC und Bewertung analytischer Methodentransfers

Von dem Fachbereich für Chemie und Pharmazie
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr.rer.nat.)
genehmigte
D i s s e r t a t i o n

von
Udo Schepers
aus Papenburg

1. Referentin oder Referent:	Prof. Dr. H. Wätzig
2. Referentin oder Referent:	Prof. Dr. P. Mischnick
eingereicht am:	12.12.2005
mündliche Prüfung (Disputation) am:	28.02.2006

2006

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung des Fachbereichs für Chemie und Pharmazie, vertreten durch Prof. Dr. H. Wätzig in folgenden Beiträgen vorab veröffentlicht:

Publikationen

U. Schepers, J. Ermer, L. Preu, H. Wätzig

Wide concentration range investigation of recovery, precision and error structure in liquid chromatography

Journal of Chromatography B, 810 (2004) 111–118

U. Schepers, H. Wätzig

Application of the equivalence test according to a concept for analytical method transfers from the International Society for Pharmaceutical Engineering (ISPE)

Journal of Pharmaceutical and Biomedical Analysis 39 (2005) 310–314

U. Schepers, H. Wätzig

Application of the equivalence test for analytical method transfers: testing precision using the United States Pharmacopoeia concept <1010>

Journal of Pharmaceutical and Biomedical Analysis 40 (2006) (accepted for publication)

Tagungsbeiträge

U. Schepers, H. Wätzig

Poster: Analytischer Methodentransfer - Möglichkeiten der Datenauswertung

Jahrestagung der Deutschen Pharmazeutischen Gesellschaft, Berlin (2002)

U. Schepers, H. Wätzig

Vortrag: Äquivalenztest oder klassischer t-Test? Beurteilung der Leistungsfähigkeit für den Bereich Methodentransfer

Jahrestagung der Fachgruppe Arzneimittelkontrolle und Pharmazeutische Analytik, Mainz (2005)

Danksagung

Die vorliegende Arbeit wurde am Institut für Pharmazeutische Chemie der Technischen Universität Carolo-Wilhelmina zu Braunschweig unter Anleitung von Herrn Prof. H. Wätzig angefertigt, dem ich an dieser Stelle ganz herzlich für die Bereitstellung und die Unterstützung bei der Bearbeitung des Themas danken möchte.

Ein weiterer Dank gilt Frau Prof. P. Mischnick für die Übernahme des Korreferats.

Den Kollegen aus dem Arbeitskreis danke ich für die gute Zusammenarbeit, besonders Heidi Köppel, Michael Graf und Simone Schröder für die fachliche Unterstützung, Diskussionen sowie für das Korrekturlesen.

Dr. Johann Grünefeld und Hendrik Stukenbrock möchte ich für die gute Zusammenarbeit bei der Semesterbetreuung danken.

Weiterhin danke ich allen anderen Mitarbeitern des Instituts für das gute Arbeitsklima.

Für die Realisierung der HPLC-Gehaltsbestimmungen im Rahmen eines Drittmittelprojektes mit Sanofi-Aventis danke ich Dr. Joachim Ermer, sowie Dr. Lutz Preu für die fachliche Unterstützung und Brigitte Babic für die Mitarbeit bei den Probenaufbereitungen und Messungen.

für Jutta

Inhaltsverzeichnis

1	Einleitung und Aufgabenstellung	1
2	Theoretischer Teil	4
2.1	Hochleistungs-Flüssigchromatographie (HPLC)	4
2.1.1	Prinzip und Aufbau	4
2.1.2	Signal-Rausch-Verhältnis	6
2.1.3	Auflösung	7
2.2	Präzision	8
2.2.1	Definition	8
2.2.2	Berechnung der Präzision	8
2.2.3	Präzisionsebenen	9
2.3	Analytischer Methodentransfer in der Pharmazeutischen Analytik	11
2.4	Statistische Überprüfung eines Methodentransfers	12
2.5	Allgemeiner Ablauf eines statistischen Tests	13
2.5.1	Vorbemerkung zu statistischen Methoden	13
2.5.2	Aufstellen von Hypothesen	14
2.5.3	Testen der Hypothesen	15
2.5.4	Treffen einer Entscheidung	15
2.6	Statistische Testungen	17
2.6.1	Klassischer Zweistichproben-t-Test	17
2.6.2	Beurteilung des Zweistichproben-t-Testes im Rahmen des analytischen Methodentransfers	20
2.6.3	Konstruktionsprinzip von Äquivalenztests	25
2.6.4	Äquivalenztest für den Mittelwertsvergleich	27
2.6.4.1	Einsatz absoluter Akzeptanzgrenzen	28
2.6.4.2	Einsatz relativer Akzeptanzgrenzen	29
2.6.4.3	Verwendung von relativen Akzeptanzgrenzen und eines Konfidenzintervalls	30
2.6.5	Vorteil des Äquivalenztestes gegenüber dem Zweistichproben-t-Test	32
2.6.6	Nichtstatistischer Vergleich der Mittelwerte	35
2.6.7	Klassischer F-Test	38
2.6.8	Nachteil des F-Testes	39
2.6.9	Äquivalenztest für den Vergleich zweier Varianzen	41

2.6.10 Vorteil des Äquivalenztestes im Vergleich zum F-Test	43
2.6.11 Testung der Standardabweichung gegen eine Akzeptanzgrenze	45
3 Präzision und Wiederfindung in der HPLC	46
3.1 Methodenoptimierung	46
3.1.1 Verkürzung der Retentionszeit	47
3.1.2 Erhöhung der Stabilität der Probenlösung	47
3.1.3 Säulenofen	49
3.1.4 Verringerung der Peakbreite bei Injektion von 100 µL (Reihe B)	49
3.1.5 Systemeignungstest	51
3.2 Ergebnisse zur Präzision	52
3.2.1 Beziehung zwischen der Präzision und der Konzentration	54
3.2.1.1 Ergebnisse der Reihe A	54
3.2.1.2 Ergebnisse der Reihe B	62
3.2.1.3 Ergebnisse der Reihe C	65
3.2.2 Beziehung zwischen der Systempräzision und der Detektionswellenlänge	69
3.2.3 Relative Zusammensetzung der Wiederhol-Standardabweichung	71
3.2.4 Abschätzung der Mehrtages-Standardabweichung	77
3.3 Ergebnisse zur Wiederfindung	79
3.3.1 Ergebnisse der Reihe A	79
3.3.2 Ergebnisse der Reihe C	83
3.4 Vergleich von statistischem und empirischem Vertrauensbereich anhand der Mindestanforderungen der ICH-Richtlinie Q2B	86
4 Leistung von Äquivalenztests beim analytischen Methodentransfer	89
4.1 Notwendige Voraussetzungen	89
4.2 Leistung beim Vergleich von Mittelwerten	89
4.2.1 Vereinfachung bezüglich der Varianzen	89
4.2.2 Designauswahl	90
4.2.3 Leistungsfähigkeit des Äquivalenztestes ausgehend vom ISPE Versuchsdesign	90
4.2.4 Ergebnisse	92
4.2.5 Ergebnisse bei Verwendung verschiedener Chargen	96
4.2.6 Abhängigkeit der Power vom Bias	97

4.2.7	Leistungsfähigkeit des Äquivalenztestes unabhängig von der Akzeptanzgrenze (AkzG)	99
4.3	Leistung beim Vergleich von Varianzen	103
4.3.1	Voraussetzung	103
4.3.2	Auswahl der Testung	103
4.3.3	Leistungsfähigkeit des Äquivalenztestes	103
4.3.4	Ergebnisse für die USP <1010> Methode	104
4.4	Testung der Standardabweichung gegen eine Akzeptanzgrenze	107
5	Zusammenfassung	109
6	HPLC-experimenteller Teil	113
6.1	Geräte	113
6.2	Analyten	114
6.3	Reagenzien	116
6.4	Methoden	116
6.4.1	Reihe A	116
6.4.2	Reihe B	117
6.4.3	Reihe C	117
6.5	Messserien	117
6.6	Injektionssequenzen	117
6.7	Probenaufbereitung	118
6.7.1	Mobile Phase	118
6.7.2	Lösungsmittel	118
6.7.3	Tablettenmatrix	118
6.7.4	Einwaage der Proben	119
6.7.5	Standards	119
6.7.6	Proben der Reihe A	119
6.7.7	Proben der Reihe B	120
6.7.8	Proben der Reihe C	120
6.8	Datenauswertung	122
7	Simulationen und Berechnungen	123
7.1	Simulation des ISPE-Versuchsdesigns	123
7.1.1	Allgemeine Durchführung einer Simulation	123
7.1.2	Aufbau der Grundgesamtheiten	125
7.1.3	Berechnung der Grundgesamtheiten	127

7.1.4	Validierung der Tests	127
7.1.5	Kompromiss zwischen Leistungsfähigkeit und Präzision	128
7.1.6	Einfluss der Varianzkomponenten	128
7.1.7	Einfluss der Verringerung der $\hat{\sigma}_{sp}$ oder der $\hat{\sigma}_{sys}$ auf die $\hat{\sigma}_{\bar{x}_{an}}$	129
7.1.8	Simultane Testung dreier verschiedener Chargen anhand eines globalen Signifikanzniveaus	130
7.1.9	Einsatz des verbundenen t-Tests	131
7.2	Berechnungen zur Testung nach USP <1010>	132
7.2.1	Methode	132
7.2.2	Berechnung der Power	132
7.2.3	Berechnung der Hälfte der oberen Akzeptanzgrenze	132
8	Abkürzungsverzeichnis	134
9	Literaturverzeichnis	136
10	Anhang	141
10.1	Exemplarisches Probenaufbereitungsschema aus Reihe C	141
10.2	Tabellen der Einzelwerte aus der Messreihen A	142
11	Lebenslauf	149

1 Einleitung und Aufgabenstellung

Die HPLC ist die am häufigsten eingesetzte Analysetechnik zur Quantifizierung von Arzneistoffen in der Pharmazeutischen Analytik [1, 2]. Zum ersten Mal angewandt wurde das chromatographische Trennprinzip vom russischen Botaniker Mikhail S. Tsvet im Rahmen seiner Forschung an Chlorophyll im Jahre 1903 [3]. Einen großen Schritt in Richtung der modernen HPLC schafften A. J. P. Martin und R. L. M. Synge in den Fünfziger Jahren mit theoretischen Überlegungen, dass für eine wirksame Chromatographie in flüssiger Phase sehr kleine Teilchen und hoher Druck notwendig sind, wofür sie 1952 den Nobelpreis erhielten [4]. Die praktische Umsetzung der HPLC gelang Ende der Sechziger Jahre durch Einführung von Teilchen mit Durchmessern unter 10 µm sowie einer neuen Generation von Gerätemodulen. Der Durchbruch erfolgte nach der Entwicklung von sogenannten Reversed-Phase-Säulen Ende der Siebziger Jahre, die heute 75 % aller Trennungen ausmachen [5].

Analytische Methoden - und damit insbesondere Methoden der HPLC - werden in der Pharmazeutischen Analytik überall dort eingesetzt, wo Arzneistoffe oder -formen entwickelt oder produziert werden. Aufgaben in der Qualitätskontrolle wie die Freigabe von Chargen aufgrund von Gehalts- und Reinheitsbestimmungen oder Aussagen zur Haltbarkeit durch Stabilitätsstudien basieren auf analytischen Ergebnissen [6]. Nach dem Arzneimittelgesetz AMG sowie aufgrund zahlreicher GMP-Richtlinien z.B. der ICH-Richtlinie Q7A [7] (ICH = International Conference on Harmonisation) ist der pharmazeutische Unternehmer zur Validierung seiner Herstellungsverfahren und damit auch der eingesetzten Analytik verpflichtet [8].

In den ICH-Richtlinien Q2A [9] und Q2B [10] zur Validierung analytischer Methoden werden Untersuchungen zur Präzision und Richtigkeit gefordert. Da zudem die durch Präzision und Richtigkeit geprägte analytische Unsicherheit von Methoden zur Gehaltsbestimmung einen großen Einfluss auf das Setzen von Akzeptanzgrenzen hat, besteht ein starkes Interesse an technisch erreichbaren Präzisionen in der HPLC [11, 12].

Für die Wiederhol-Standardabweichung einiger HPLC-Methoden werden in [13] anhand der Auswertung von Systemeignungstests Werte von 0,5 bis 1,0 % als

RSD% gefunden. Eine ausführliche Übersichtarbeit zur Ermittlung der Wiederhol-Standardabweichung HPLC-analytischer Methoden wurde in [14] anhand von 156 Stabilitätsstudien erstellt. Darin konnte für die Arzneiform Tablette ein mittlerer RSD%-Wert von 0,80 % abgeleitet werden ($n = 90$). Mit 0,60 % vergleichbare Werte für die Wiederhol-Standardabweichung lieferten Ringversuche [11, 15]. Weitere Arbeiten klären und bemaßen einzelne Variationskomponenten der Wiederhol-Standardabweichung [16, 17]. Die zahlreich zu findenden Wiederhol-Standardabweichungen einzelner neu entwickelter Methoden sind oft aus nur wenigen Serien abgeschätzt worden.

Bisher wurde noch keine zusammenhängende Studie einer HPLC-Methode zur Präzision und Wiederfindung über einen weiten Konzentrationsbereich durchgeführt, bei der dieser Bereich schrittweise in kleinen Intervallen abgedeckt wurde. Obwohl Probenaufbereitung und Integration oft als Hauptfehlerquellen diskutiert worden sind [16, 18], ist ihr quantitativer, sich stetig wandelnder Beitrag in Abhängigkeit vom Gesamtfehler nicht offenkundig. Des Weiteren stellt sich die Frage, von welchen Konzentrationen an der Integrationsfehler die dominierende Rolle spielt.

Ein Ziel dieser Arbeit ist es, diese Fragestellungen mit einer repräsentativen und robusten HPLC-Methode, mit einem stabilen Arzneistoff sowie einer Reihe gut charakterisierter, möglicher Verunreinigungen zu untersuchen.

Präzision und Wiederfindung spielen nicht nur bei der Validierung analytischer Methoden, sondern auch im Bereich des Transfers analytischer Methoden eine tragende Rolle [19]. Ein analytischer Methodentransfer soll die Zieleinheit in die Lage versetzen, ein validiertes Prüfverfahren aus der Ursprungseinheit für den Einsatz in der Routineanalytik vor Ort richtig und reproduzierbar durchführen zu können [20]. Über den Erfolg eines analytischen Methodentransfer entscheidet die Datenauswertung.

Die Vorgehensweise zur Datenauswertung ist von den Behörden nicht explizit vorgeschrieben. Es existieren eine Reihe von Vorschlägen hinsichtlich der statistischen Testungen sowie des Versuchsdesigns [19-22]. Es soll untersucht werden, ob der Äquivalenztest dieser speziellen Testsituation beim

Methodentransfer am besten gerecht wird, wenn man sich an den mit der Entscheidung über Annahme oder Ablehnung eines Transfers verbundenen Risiken orientiert.

Ein weiteres Ziel dieser Arbeit ist es daher, die Leistungsfähigkeit des Äquivalenztestes für den Vergleich von Mittelwerten (Richtigkeit) und Varianzen (Präzision) der beteiligten Labors anhand von Simulationen und Berechnungen zu untersuchen.

2 Theoretischer Teil

2.1 Hochleistungs-Flüssigchromatographie (HPLC)

2.1.1 Prinzip und Aufbau

Das Prinzip der Chromatographie beruht auf dem Übergang der zu trennenden Komponenten zwischen zwei nicht mischbaren Phasen. In der Flüssigchromatographie (s. Abb. 2.1-1) wird dazu die Probe in der mobilen Phase gelöst und über eine stationäre Phase bewegt, die sich entweder in einer Säule oder auf einem Flachbett befindet. Aufgrund der Wechselwirkungen der Komponenten mit der stationären Phase kommt es bei genügend langer Laufzeit zu deren Trennung [5].

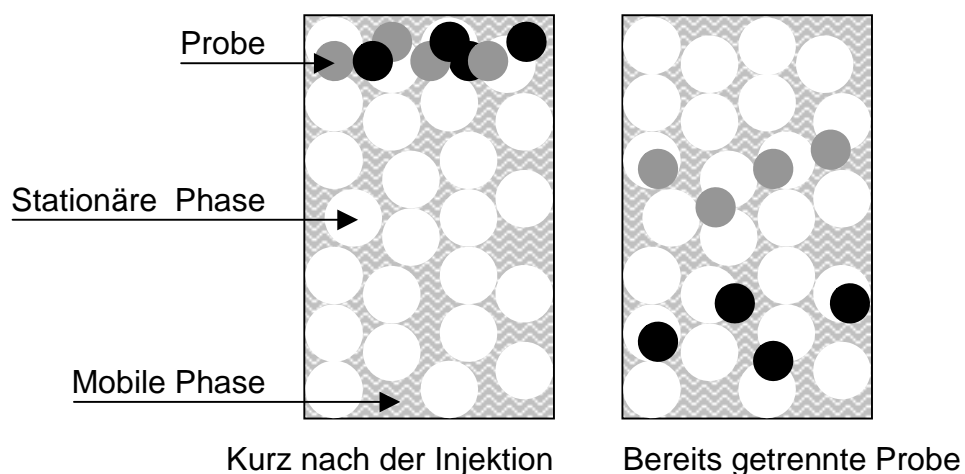


Abbildung 2.1-1: Chromatographischer Trennprozess.

Bei der Hochleistungs-Flüssigchromatographie (High Performance Liquid Chromatography, HPLC) wird die mobile Phase unter hohem Druck durch eine Trennsäule gepresst. Dazu ist eine HPLC-Anlage erforderlich, deren Zusammensetzung aus verschiedenen Komponenten in Abbildung 2.1-2 schematisch dargestellt ist.

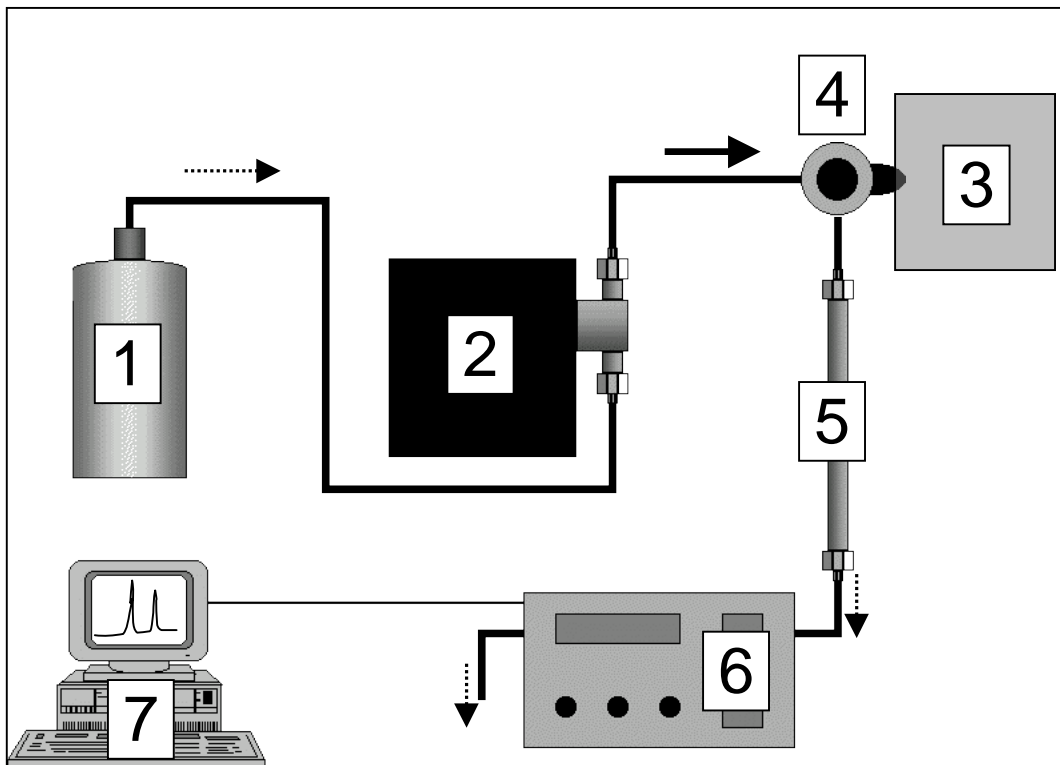


Abbildung 2.1-2: Schematischer Aufbau einer HPLC-Anlage.

Die mobile Phase wird dabei aus einem Vorratsgefäß (1) über eine Zuleitung von der Hochdruckpumpe (2) angesaugt und unter einem Druck von bis zu 400 bar weiter zur Trennsäule (5) geleitet. Die zu analysierende Probe wird von einem Probenwechsler (Autosampler) (3) über ein Injektionsventil (4) zur mobilen Phase gegeben. Beim Durchfließen der Trennsäule (5) - eine Standardsäule ist 250 mm lang und hat einen Innendurchmesser von 4,6 mm - erfolgt die Auftrennung der Probe in die Komponenten. Anschließend werden die Komponenten zum Detektor (6) weiterbefördert. Dort werden Signale registriert, die von einem Rechner (7) zu einem Chromatogramm verarbeitet werden.

2.1.2 Signal-Rausch-Verhältnis

Das Signal-Rausch-Verhältnis (S/N) diene bei den HPLC-Bestimmungen in dieser Arbeit als Maß, um festzustellen, welche Verdünnungsschritte noch vermessen werden konnten. Das Signal-Rausch-Verhältnis beeinflusst die Präzision der Bestimmung und wird nach Gleichung 2.1-1 berechnet [23]:

$$S / N = \frac{2H}{h} \quad (2.1-1)$$

H = Höhe des Peaks der betreffenden Komponente im Chromatogramm, gemessen vom Peakmaximum bis zur virtuellen Basislinie des Signals.

h = Höhe des gesamten Untergrundrauschens unter- und oberhalb der Basislinie in einem Chromatogramm einer Blindlösung betrachtet über eine Distanz symmetrisch um den Peak, die dem 20-fachen der halben Peakbreite entspricht (s. Abb. 2.1-3).

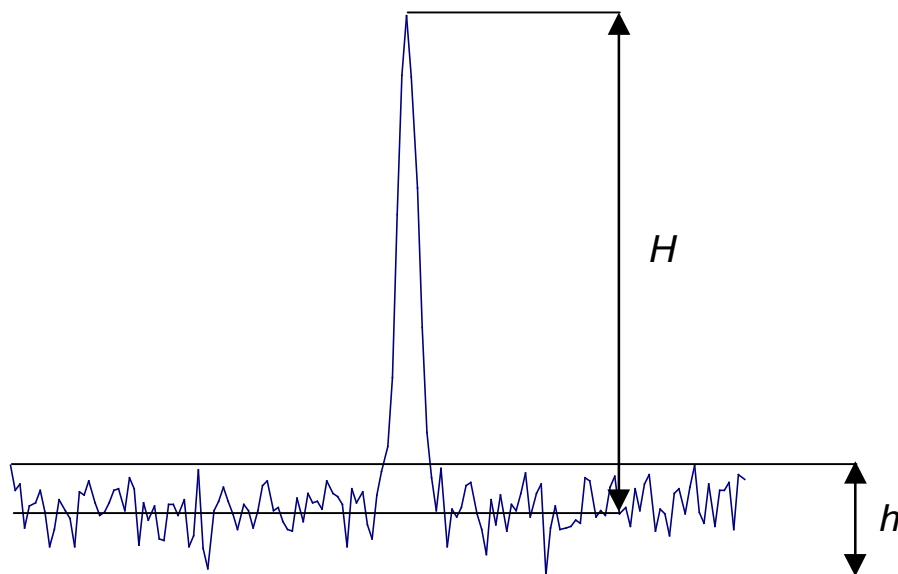


Abbildung 2.1-3: Peakhöhe H und Höhe des Untergrundrauschens h in einem Beispielchromatogramm.

2.1.3 Auflösung

Einige HPLC-Bestimmungen in dieser Arbeit (Reihe C) beinhalteten die Quantifizierung mehrerer Komponenten in einem Chromatogramm. Bei der Quantifizierung mehrerer Peaks spielt die zwischen benachbarten Peaks vorhandene Auflösung R eine große Rolle. Eine unzureichende Auflösung übt einen verzerrenden Effekt auf die absolute Peakfläche beider Peaks und auf die Präzision bei Wiederholbestimmungen aus.

Die praktische Bestimmung der Auflösung zwischen zwei Peaks erfolgt durch Bildung des Quotienten aus dem Abstand der beiden Peakmaxima voneinander und dem arithmetischen Mittel aus den beiden Basisbreiten (Gl. 2.1-2) [24]:

$$R = 2 \frac{t_{R2} - t_{R1}}{w_1 + w_2} = 1,18 \frac{t_{R2} - t_{R1}}{w_{\frac{1}{2}1} + w_{\frac{1}{2}2}} \quad (2.1-2)$$

Beim zweiten Term sind im Nenner die Peakbreiten auf halber Höhe gemeint. Die Differenz im Zähler und die Summe im Nenner werden in der selben Einheit aus dem Chromatogramm abgelesen, daher ist die Auflösung dimensionslos. Bei einer Auflösung von 1 sind die Peaks erkennbar, aber nicht vollständig voneinander getrennt. Von einer Basislinientrennung zwischen zwei gleichgroßen Peaks kann man bei einer Auflösung von 1,5 ausgehen. Je asymmetrischer die beiden Peaks, desto höher sollte die Auflösung über 1,5 liegen [24].

2.2 Präzision

2.2.1 Definition

Die Präzision einer Analysenmethode ist das Maß für die Übereinstimmung zwischen den Messwerten einer Serie, die aus mehreren Stichproben aus dem gleichen homogenen Probenmaterial unter vorgeschriebenen Bedingungen erhalten wurde [9].

2.2.2 Berechnung der Präzision

Die Präzision einer Analysenmethode wird üblicherweise als Varianz $\hat{\sigma}^2$, Standardabweichung $\hat{\sigma}$ oder als prozentuale relative Standardabweichung RSD% angegeben [25]. Ausgehend von einer Stichprobe wird die Varianz einer Grundgesamtheit nach Gleichung 2.2-1 geschätzt:

$$\hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} \quad (2.2-1)$$

Wobei der Term x_i den i-ten Einzelwert, \bar{x} den arithmetischen Mittelwert und n die Stichprobengröße darstellen. Bei Berechnungen statistischer Tests wird häufig die Standardabweichung $\hat{\sigma}$ herangezogen. Sie ist ein absolutes Maß, deren Größe allein nichts über die Qualität der Präzision aussagt. Die Standardabweichung ergibt sich nach Gleichung 2.2-2 als Wurzel der Varianz zu:

$$\hat{\sigma} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (2.2-2)$$

Die prozentuale relative Standardabweichung RSD% ist die gängigste Angabe der Präzision. Sie eignet sich besonders zum Vergleich verschiedener Streuungen, da sie die Abweichungen relativ zur Größe des Mittelwertes \bar{x} angibt. Dazu wird die Standardabweichung auf den Mittelwert der Stichprobe bezogen:

$$\text{RSD\%} = \frac{\hat{\sigma}}{\bar{x}} \cdot 100 \quad (2.2-3)$$

2.2.3 Präzisionsebenen

Betrachtet man eine analytische Methode, so wird jeder bei Anwendung der Methode angewandte Schritt einen gewissen Anteil zur gesamten Variation beitragen. In der Analytik werden daher verschiedene Präzisionsebenen unterschieden, je nachdem, welche Einflussgrößen in die betrachtete Präzision mit einfließen [10, 26]:

Die Systempräzision $\hat{\sigma}_{\text{sys}}$ (system precision) wird durch die wiederholte Messung derselben Probe erhalten. Sie wird im Wesentlichen vom Messinstrument beeinflusst. Bei Quantifizierungen niedriger Konzentrationen anhand von Peakflächen kann auch die Integration in einem hohen Maße zur Systempräzision beitragen.

Die Wiederhol-Standardabweichung $\hat{\sigma}_r$ (Wiederholpräzision, repeatability) wird durch die mehrfache Anwendung des gesamten Analyseverfahrens inklusive Probenvorbereitung in einem Labor vom gleichen Operator mit demselben Messinstrument auf identisches Testmaterial erhalten. Dabei werden die Messungen innerhalb eines kurzen Zeitraumes nacheinander durchgeführt (innerhalb einer Serie, mit den selben Referenzstandards).

Die Mehrtages-Standardabweichung $\hat{\sigma}_{\text{int}}$ (intermediate precision) gibt die Streuung an, wenn die gleiche, vollständige Methode an verschiedenen Tagen, möglicherweise durch unterschiedliche Analytiker und mit unterschiedlicher Ausrüstung, aber noch innerhalb des gleichen Labors durchgeführt wird. Zusätzliche Streuungen des Standards (bei Kalibrierungen) sowie Einflüsse durch zeitliche Veränderungen (z.B. Temperaturschwankungen) werden miterfasst.

Die Ringversuchs-Standardabweichung $\hat{\sigma}_R$ (reproducibility) wird mit identischer Methode und identischem Testmaterial in unterschiedlichen Labors ermittelt. Bei diesem Präzisionsmaß werden besonders auch systematische Veränderungen durch äußere Einflussgrößen (z.B. Temperatur, Raumfeuchte, Reagenzien) berücksichtigt. Als grobe Abschätzung einer zu erwartenden $\hat{\sigma}_R$ kann ein Verhältnis von $\hat{\sigma}_R = 1,7 \cdot \hat{\sigma}_r$ herangezogen werden [27].

Die nachfolgende Abbildung 2.2-1 gibt eine Übersicht über die innerhalb eines Labors betrachtaren Präzisionsebenen sowie wichtiger Einflussgrößen.

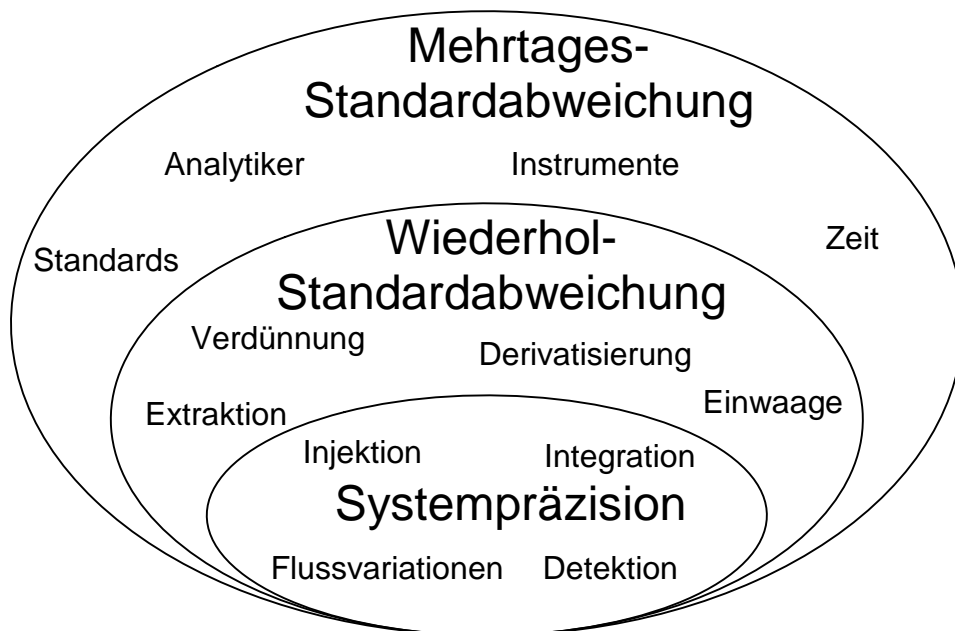


Abbildung 2.2-1: Graphische Darstellung der innerhalb eines Labors betrachtaren Präzisionsebenen. Die jeweils tiefere Ebene ist in der hierarchisch höheren enthalten (aus [21] modifiziert übernommen).

2.3 Analytischer Methodentransfer in der Pharmazeutischen Analytik

Der Transfer analytischer Methoden ist ein wichtiger Bestandteil der täglichen Arbeit von Labors in der Pharmazeutischen Industrie. In den vergangenen Jahren hat sich die Sichtweise durchgesetzt, dass der Nutzen von formal durchgeführten Methodentransfers für die Routinekontrolle, die für ein erfolgreiches Produkt über einen langen Zeitraum von zehn bis zwanzig Jahre gehen kann, den Aufwand rechtfertigt [21].

Ein analytischer Methodentransfer soll die Zieleinheit in die Lage versetzen, ein validiertes Prüfverfahren aus der Ursprungseinheit für den Einsatz vor Ort richtig und reproduzierbar durchführen zu können [20]. So kann der Transfer eines validierten Prüfverfahrens aus der Ursprungseinheit z.B. in ein Labor der Qualitätskontrolle der Produktionseinheit, an Tochterfirmen, an einen Lizenznehmer, an einen Auftragshersteller oder an ein Auftragslabor erfolgen.

Die allgemeine Durchführung eines Methodentransfers ist ein mehrstufiger Prozess, dessen zentraler Bestandteil die Erstellung eines Transferstrategiedokumentes ist, welches alle zu transferierenden analytischen Verfahren und Maßnahmen enthält, die zum erfolgreichen Transfer notwendig sind.

Letztlich wird ein Transfer durch die Erstellung und Genehmigung eines Transferberichts abgeschlossen, aus dem auch die Art und Weise der Datenauswertung hervorgeht.

Ein Transfer nimmt üblicherweise mehrere Wochen in Anspruch und sollte im Falle des Transfers in ein Labor der Qualitätskontrolle für ein neues Produkt ungefähr 18 Monate vor der routinemäßigen Anwendung beginnen, um dem Empfängerlabor ausreichende Gelegenheit zur Rückkopplung und zur Realisierung möglicher Investitionen zu geben [21].

2.4 Statistische Überprüfung eines Methodentransfers

Die Durchführung des formellen Transfers im Rahmen eines abschließenden Ringversuchs zwischen den betreffenden Labors liefert die Daten entsprechend dem vorher festgelegten Versuchsdesign. Die Entscheidung darüber, ob der Transfer geglückt ist, erfolgt anhand der Auswertung dieser Daten gegenüber vorgegebenen oder selbstentwickelten Akzeptanzgrenzen.

Das Versuchsdesign richtet sich nach der Art der zu testenden Methode. In dieser Arbeit wird ausschließlich die Gehaltsbestimmung betrachtet. Im Rahmen der Testung eines Produktes werden darüber hinaus noch weitere Methoden z.B. zur Untersuchung der Freisetzung oder zur Quantifizierung von Verunreinigungen und Nebenprodukten transferiert. Diese bedingen andere Versuchsdesigns und andere Akzeptanzgrenzen.

Im Rahmen von Methoden zur Gehaltsbestimmung werden die Parameter Wiederfindung und Präzision anhand von Mittelwert und Standardabweichung untersucht.

Im Prinzip sind drei Vorgehensweisen zur Testung der Mittelwerte und Varianzen der Labors möglich [21, 28]:

- Anwendung von Äquivalenztests
- Anwendung von Zweistichproben-t-Test und F-Test
- Nichtstatistischer Vergleich gegen absolute Akzeptanzgrenzen

Die Vor- und Nachteile dieser Testungen werden in den Abschnitten 2.6.1 bis 2.6.11 näher erläutert.

Die Akzeptanzgrenzen, innerhalb derer ein Transfer als erfolgreich beurteilt wird, können anhand statistischer Überlegungen hergeleitet werden. Die Spezifikationsgrenzen und die produktionsbedingte Variabilität sollten mit einbezogen werden [6, 29], um in der Qualitätskontrolle OOS-Ergebnisse (Out Of Specification) zu vermeiden. Der ISPE-Guide empfiehlt für Spezifikationsgrenzen von 95 bis 105 % (häufigster Fall) analytische Akzeptanzgrenzen von $\pm 2\%$ [19].

2.5 Allgemeiner Ablauf eines statistischen Tests

2.5.1 Vorbemerkung zu statistischen Methoden

Bei analytischen Daten handelt es sich um Stichproben aus einer größeren, wirklichen oder hypothetischen Population. Aus dieser Stichprobe sollen Erkenntnisse über die gesamte Population (z.B. den Gehalt einer Charge) gewonnen werden. Dabei muss man die zufälligen Schwankungen der einzelnen Messwerte um den entsprechenden wahren Wert in der Gesamtpopulation berücksichtigen und quantitativ erfassen. Hierzu dient die schließende Statistik. Sie baut auf Modellen und Gesetzen über den Zufall auf, die in der Wahrscheinlichkeitstheorie hergeleitet worden sind. Dabei versucht man, möglichst einfache und bewährte statistische Modelle an die Daten anzupassen (Schätzungen) und die Güte der Anpassung zu prüfen (Tests).

Die hier im Folgenden eingesetzten statistischen Tests basieren alle auf der Normalverteilung $N(\mu, \sigma^2)$. Die Normalverteilung besitzt in der Statistik eine große Bedeutung. Die der Normalverteilung zugrundeliegende Zufallsvariable x besitzt den Erwartungswert μ und die Varianz σ^2 [30]. Die Dichtefunktion der Normalverteilung weist eine glockenförmige Gestalt auf und ist symmetrisch um μ (Gl. 2.5-1).

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.5-1)$$

Abbildung 2.5-1 zeigt die Dichtefunktion der Standardnormalverteilung $N(0,1)$ mit dem Erwartungswert 0, der Varianz 1 und den Wendepunkten an den Stellen $\mu \pm \sigma$ (-1,1).

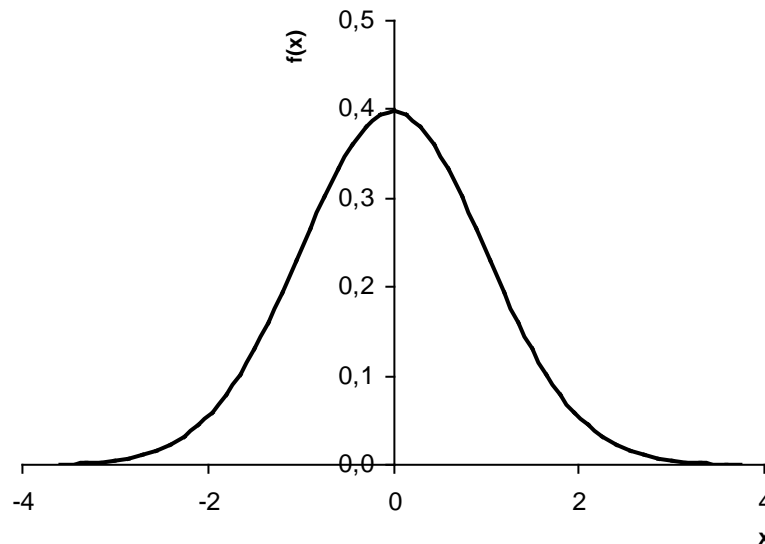


Abbildung 2.5-1: Dichtefunktion der Standardnormalverteilung $N(0,1)$.

Die große Bedeutung der Normalverteilung folgt aus dem Zentralen Grenzwertsatz. Dieser erlaubt es, eine Zufallsvariable, die sich additiv aus unabhängigen Einzelwirkungen zusammensetzt, als angenähert normalverteilt anzunehmen [25]. Da ein analytisches Ergebnis eine solche zusammengesetzte Zufallsvariable darstellt, kann von einer Normalverteilung analytischer Messwerte ausgegangen werden.

2.5.2 Aufstellen von Hypothesen

Die Durchführung eines statistischen Tests beginnt mit der Formulierung je einer Null- und Alternativhypothese (H_0 und H_1). „Unter einer Hypothese versteht man in der Statistik eine Annahme über die Verteilung einer Zufallsvariablen“ [31]. Eine Hypothese kann z.B. die Annahme sein, dass die betreffende Verteilung den Mittelwert 100,3 hat. Die Alternativhypothese bestünde dann beispielsweise in der Annahme, dass sich der Mittelwert von 100,3 unterscheidet (sog. zweiseitiger Test). Die beiden Hypothesen H_0 und H_1 beschreiben damit die Fragestellung, die

mit dem Test beantwortet werden soll, z.B. die Verteilung hat den Mittelwert 100,3 ($H_0: \mu = 100,3$) bzw. der Mittelwert der Verteilung ist ungleich 100,3 ($H_1: \mu \neq 100,3$). Null- und Alternativhypothese bilden das logische Gegenteil zueinander und ergeben insgesamt alle möglichen Ereignisse.

2.5.3 Testen der Hypothesen

Nach der Formulierung beider Hypothesen erfolgt die Überprüfung der Nullhypothese anhand eines statistischen Tests [25]. Ein solcher Test prüft, ob ein Parameter θ (beim Methodentransfer die Differenz der Mittelwerte $\theta = \mu_1 - \mu_2$ (s. 2.6.4) oder der Quotient zweier Varianzen $\theta = \hat{\sigma}_2^2 / \hat{\sigma}_1^2$ (s. 2.6.7)) größer oder kleiner als ein bestimmter Sollwert θ_0 ist. In so einem Fall ordnet man mit Hilfe einer sogenannten Teststatistik (Prüfgröße) T den Beobachtungen des Experimentes x_1, \dots, x_n eine Zahl $T(x_1, \dots, x_n)$ zu, bei der ein großer Wert eher für das Vorliegen von H_1 und ein kleiner eher für das von H_0 spricht. Außerdem sucht man eine Zahl $c_{1-\alpha}$ (sog. kritischer Wert), für die gilt:

$$P_{\theta}(T > c_{1-\alpha}) \leq \alpha \quad \text{für alle } \theta \in H_0,$$

(P_{θ} : Wahrscheinlichkeit beim Vorliegen von θ).

2.5.4 Treffen einer Entscheidung

Der Vergleich der Prüfgröße T mit dem kritischen Wert $c_{1-\alpha}$ führt entweder zur Annahme oder zur Verwerfung von H_0 , je nachdem, ob die berechnete Prüfgröße größer oder kleiner als der kritische Wert ist. Trifft man die Entscheidung für H_0 , falls die Teststatistik kleiner als der kritische Wert ist, und für H_1 falls $T > c_{1-\alpha}$, dann handelt es sich um einen Test zum Niveau α . Da die Teststatistik nur auf Wahrscheinlichkeitsmodellen basiert, sind in Abhängigkeit von der unbekannten Wirklichkeit, nicht zwei, sondern vier Ereignisse möglich. Diese Ereignisse sind in folgender Tabelle 2.5-1 dargestellt:

In Wirklichkeit ist:

Aufgrund des statistischen Testes wird:	H_0 korrekt	H_1 korrekt
H_0 angenommen	richtige Entscheidung	Fehler 2. Art β -Fehler
H_1 angenommen	Fehler 1. Art α -Fehler	richtige Entscheidung Power ($1-\beta$)

Tabelle 2.5-1: Alle möglichen Ereignisse nach dem Treffen einer Entscheidung.

Neben den zwei richtigen Entscheidungen tritt ein Fehler 1. Art (α -Fehler) ein, wenn aufgrund der Berechnung der Teststatistik die Nullhypothese ablehnt wird, obwohl sie hätte angenommen werden müssen (Ablehnung einer richtigen Hypothese). Einen Fehler 2. Art (β -Fehler) hingegen macht man, wenn man die Nullhypothese unberechtigt beibehält, obwohl sie hätte abgelehnt werden müssen (Nichtablehnung einer falschen Hypothese).

Die Wahrscheinlichkeit, die Alternativhypothese anzunehmen, wenn sie tatsächlich richtig ist, wird als Power bezeichnet und beträgt $1-\beta$. Mit anderen Worten: Die Power eines Testes ist die Wahrscheinlichkeit H_0 abzulehnen, wenn tatsächlich H_1 richtig ist [32].

Zu beachten ist, dass innerhalb eines Testes die Fehler 1. und 2. Art miteinander in Verbindung stehen. Sind Stichprobenumfang n und Irrtumswahrscheinlichkeit α vorgegebenem, wird der β -Fehler um so größer, je kleiner das gewählte α ist. Die Wahl des Wertes für α stellt damit einen Kompromiss zwischen beiden Fehlerarten dar. In der Statistik sind für α die Werte 0,05 und 0,01 üblich. Nur wenn n unbeschränkt wachsen darf, können α und β beliebig klein gewählt werden [25]. Allerdings kann ein hoher Wert für n in der Praxis mit einem erheblichen Aufwand verbunden sein.

2.6 Statistische Testungen

2.6.1 Klassischer Zweistichproben-t-Test

Da die Prüfgröße für den Äquivalenztest beim Mittelwertsvergleich auf dem t-Test basiert, soll dieser hier kurz vorgestellt werden. Zum Vergleich der Mittelwerte zweier unabhängiger Messreihen ist der sogenannte Zweistichproben-t-Test wohl der bekannteste Vertreter [25, 32-34]. Für den Zweistichproben-t-Test existieren mittlerweile einige Modifikationen [35-38], die aber meist nur bei Annahme anderer Verteilungen als der Normalverteilung einen Vorteil bringen und somit hier keine Berücksichtigung finden, da hier von normalverteilten Daten ausgegangen wird (s. 4.1).

Zur Anwendung des Zweistichproben-t-Tests muss unterschieden werden, ob die Varianzen bekannt oder unbekannt und ob sie gleich oder ungleich sind, um die anzuwendende Prüfgröße T_t bestimmen zu können [39].

Da in den Simulationen im Rahmen dieser Arbeit als Grundannahme von unbekannten aber gleichen Varianzen ausgegangen wurde (s. 4.2.1), kam die entsprechende Prüfgröße T_t für den Zweistichproben-t-Test zur Anwendung.

Dabei wird die Nullhypothese auf Gleichheit der den beiden Stichproben zugrunde liegenden Mittelwerte der Grundgesamtheiten geprüft:

$$H_0: \mu_1 = \mu_2 \quad \text{gegen} \quad H_1: \mu_1 \neq \mu_2.$$

Die Prüfgröße T_t berechnet sich nach Gl. 2.6-1 und 2.6-2:

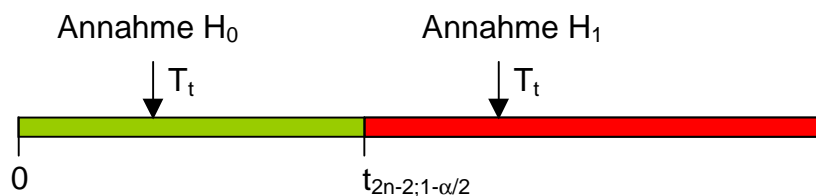
$$T_t = \frac{|\bar{x}_1 - \bar{x}_2|}{\hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2.6-1)$$

$$\hat{\sigma}_p^2 = \frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2} \quad (2.6-2)$$

Hierbei stellen \bar{x}_1 und \bar{x}_2 die geschätzten Mittelwerte, $\hat{\sigma}_1$ und $\hat{\sigma}_2$ die geschätzten Standardabweichungen und n_1 und n_2 die Stichprobenumfänge der

zu vergleichenden Messreihen dar. Da $\hat{\sigma}_1$ und $\hat{\sigma}_2$ nach Voraussetzung aus derselben Grundgesamtheit stammen (s. 4.1), können die beiden Standardabweichungen nach Gleichung 2.6-2 zu einer zusammengesetzten (engl. gepoolten) Standardabweichung $\hat{\sigma}_p$ zusammengefasst werden, so dass bei gleich großen Stichprobenumfängen $(2n - 2)$ Freiheitsgrade vorliegen.

Die Prüfgröße T_t wird mit dem kritischen Wert $t_{2n-2;1-\alpha/2}$ mit $2n - 2$ Freiheitsgraden und der Irrtumswahrscheinlichkeit α verglichen. Hierbei handelt es sich um einen sogenannten zweiseitigen Test, da sowohl auf eine Abweichung nach unten als auch nach oben geprüft wird.



Ist $T_t < t_{2n-2;1-\alpha/2}$, so wird H_0 angenommen und man unterstellt einen nicht signifikanten Unterschied. Im Fall $T_t > t_{2n-2;1-\alpha/2}$ ist die Alternativhypothese H_1 signifikant zum Niveau α abgesichert.

Beispiel:

Die Ergebnisse zweier Messreihen mit $\bar{x}_1 = 99,67$; $\bar{x}_2 = 100,14$; $\hat{\sigma}_1^2 = 0,35$; $\hat{\sigma}_2^2 = 0,21$; $n_{1,2} = 6$ und $\alpha = 0,05$ führen zu:

$$T_t = \frac{|99,67 - 100,14|}{0,53 \sqrt{\frac{1}{6} + \frac{1}{6}}} = 1,54 \qquad T_t = 1,54 < t_{10;0,975} = 2,23$$

In diesem Beispiel ist die Prüfgröße kleiner als der kritische Wert, d.h. die Nullhypothese kann zum gewählten Niveau α nicht verworfen werden.

Die Leistungsfähigkeit des Zweistichproben-t-Testes kann nicht anhand einer einzelnen Stichprobe ermittelt werden, da die wahre Abweichung zwischen den beiden Grundgesamtheiten, aus denen die Stichproben stammen, immer

unbekannt ist. Zur Ermittlung der Leistungsfähigkeit lässt sich aber für den Zweistichproben-t-Test bei vorgegebenem α die Annahmewahrscheinlichkeit in Abhängigkeit von der tatsächlichen Abweichung zwischen zwei Grundgesamtheiten (bzw. den Ergebnissen zweier Labors) darstellen (s. Abb. 2.6-1).

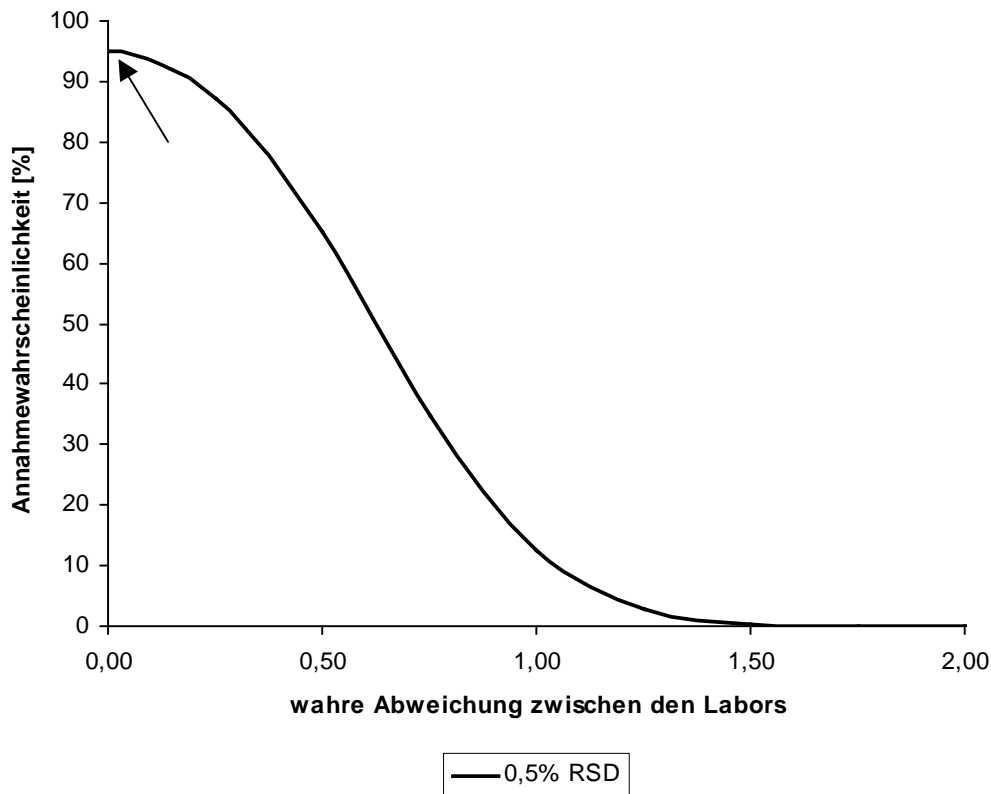


Abbildung 2.6-1: Annahmewahrscheinlichkeit des Zweistichproben-t-Testes in Abhängigkeit von einer wahren Abweichung zwischen den Labors. Die RSD% der beiden Grundgesamtheiten beträgt 0,50 % und die Stichprobengröße $n = 6$.

Abgebildet ist die Wahrscheinlichkeit, den Zweistichproben-t-Test anzunehmen, gegenüber einer wahren Abweichung zwischen den beiden Labors im Bereich von 0 bis 2 %. Die RSD% der beiden Grundgesamtheiten beträgt 0,50 % und die Stichprobengröße $n = 6$. Besteht in Wahrheit keine Abweichung zwischen den Labors (Bias = 0, s. Pfeil), so liegt die Wahrscheinlichkeit, die Alternativhypothese anzunehmen und damit einen Fehler 1. Art zu begehen, genau bei 5 % (gewählte Irrtumswahrscheinlichkeit α). In 5 von 100 Fällen erkennt der Zweistichproben-t-Test einen statistisch signifikanten Unterschied, obwohl in Wahrheit überhaupt kein Unterschied vorliegt.

In der Regel sind bei der Durchführung von Methodentransfers die Varianzen der beteiligten Labors unbekannt und wahrscheinlich ungleich. Für solche Situationen berechnet sich für den Zweistichproben-t-Test die Prüfgröße T_t nach folgenden Gleichungen (sog. Behrens-Fischer-Problem) [25]:

$$T_t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}} \quad (2.6-3)$$

$$v = \frac{\left(\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2} \right)^2}{\frac{\left(\frac{\hat{\sigma}_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{\hat{\sigma}_2^2}{n_2} \right)^2}{n_2 - 1}} \quad (2.6-4)$$

Die Freiheitsgrade für den kritischen Wert werden durch v anhand von Gleichung 2.6-4 angenähert: $t_{v; 1-\alpha/2}$. Für Situationen, bei denen $n_1 = n_2$ und $\hat{\sigma}_1 = \hat{\sigma}_2$ sind, ergibt sich $v = (2n - 2)$.

2.6.2 Beurteilung des Zweistichproben-t-Testes im Rahmen des analytischen Methodentransfers

Der Zweistichproben-t-Test hat einen großen Nachteil und ist dadurch für den Einsatz bei analytischen Methodentransfers ungeeignet:

Aus analytischer Sicht ist es sehr unwahrscheinlich, dass die Ergebnisse der Anwendung einer Methode in zwei Labors überhaupt keine Abweichung voneinander haben. Vor diesem Hintergrund erscheint die Frage: „*Sind die beiden Mittelwerte exakt gleich?*“ und die damit verbundene Nullhypothese des Zweistichproben-t-Testes auf Gleichheit ($H_0: \mu_1 - \mu_2 = 0$) ungeeignet.

Nur die Alternativhypothese wird durch das vorher festgelegte α auf 5 % abgesichert (s. Abb. 2.6-1). Der α -Fehler (s. 2.5.4) besteht in diesem Zusammenhang des Zweistichproben-t-Testes in dem Risiko, auf einen Unterschied zwischen den Labors zu schließen, obwohl in Wahrheit kein Unterschied existiert

(fälschliche Ablehnung der richtigen Nullhypothese). Damit wäre die eigentlich unnötige erneute Durchführung einer Transferstudie verbunden.

Wesentlich wichtiger aber ist es beim Methodentransfer, keinen β -Fehler zu begehen (s. 2.5.4), da dies die Annahme der Nullhypothese (die Methoden sind nicht unterschiedlich) nach sich zöge, obwohl sie falsch ist und in Wahrheit ein nicht zu akzeptierender Unterschied zwischen den Labors existiert. Die Konsequenz bestünde in der Anwendung unzureichender Analytik in der Routinefreigabe.

Beim Zweistichproben-t-Test entscheidet allein das Vorliegen (oder Nichtvorliegen) von statistischer Signifikanz über Annahme eines Transfers. Diese Eigenschaft lässt sich am einfachsten erklären, indem man den Zweistichproben-t-Test als ein um den Parameter θ berechnetes $(1-\alpha)$ -Konfidenzintervall auffasst, welches von einer unteren Grenze C_L („lower“) bzw. einer oberen Grenze C_U („upper“) umgeben ist (s. Abb. 2.6-2).

$$\begin{array}{ccc} \bar{x}_1 - \bar{x}_2 - t_{2n-2;1-\alpha/2} \cdot \hat{\sigma}_p \cdot \sqrt{2/n} = C_L & \theta = \bar{x}_1 - \bar{x}_2 & C_U = \bar{x}_1 - \bar{x}_2 + t_{2n-2;1-\alpha/2} \cdot \hat{\sigma}_p \cdot \sqrt{2/n} \\ & \downarrow & \\ & \boxed{C_L \quad \theta \quad C_U} & \end{array}$$

Abbildung 2.6-2: Graphische Darstellung des Konfidenzintervalls des Zweistichproben-t-Testes.

Die Größe dieses Konfidenzintervalls hängt von den Freiheitsgraden n und von α ab. Zusätzlich zum Konfidenzintervall werden mit $\theta_0 - \varepsilon$ und $\theta_0 + \varepsilon$ die beiden auf den Sollwert θ_0 bezogenen Akzeptanzgrenzen dargestellt, die einen analytisch relevanten Unterschied markieren sollen (s. Abb. 2.6-3 und 2.6-4).

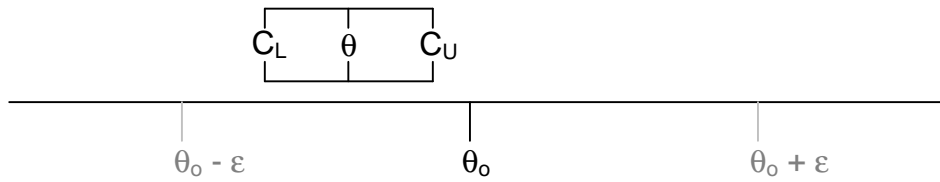


Abbildung 2.6-3: Der Sollwert θ_0 ist nicht im Konfidenzintervall enthalten. Ein signifikanter Unterschied wird geschlussfolgert. Der Abstand von θ zu θ_0 allerdings wäre akzeptabel. In diesem Fall liegt ein statistisch signifikanter aber praktisch irrelevanter Unterschied vor.

So können sich einerseits zwei Mittelwerte signifikant voneinander unterscheiden, weil sich entweder durch hohe Präzisionen oder durch eine große Stichprobenzahl eine Signifikanz ergibt, obwohl die geschätzte Differenz der Mittelwerte eigentlich akzeptabel wäre.

Andererseits ist es möglich, dass aufgrund einer geringen Präzision oder einer kleinen Stichprobenzahl eine an sich nicht akzeptable Differenz als nicht signifikant eingeordnet wird (s. Abb. 2.6-4).

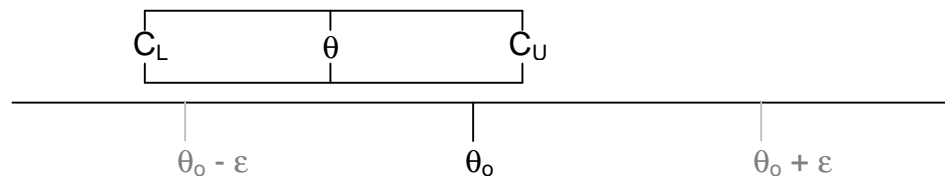


Abbildung 2.6-4: Ein nicht akzeptabler Unterschied wird übersehen.

Der Sollwert θ_0 liegt innerhalb des Konfidenzintervalls, obwohl die untere Grenze C_L des geschätzten Konfidenzintervalls um θ in einem nicht akzeptablen Abstand zum Sollwert liegt. In diesem Fall würde beim klassischen Zweistichproben-t-Test ein praktisch relevanter Unterschied übersehen, da nur die Lage des Sollwertes zum Konfidenzintervall entscheidend ist.

In welchem Maße die Parameter Präzision und Stichprobengröße die Annahmewahrscheinlichkeit des Zweistichproben-t-Testes beeinflussen, zeigt Abbildung 2.6-5. Dort werden mehrere Transfersituationen verglichen, die sich in der Präzision der beteiligten Labors voneinander unterscheiden (0,5; 1,0 und 2,0 % RSD%).

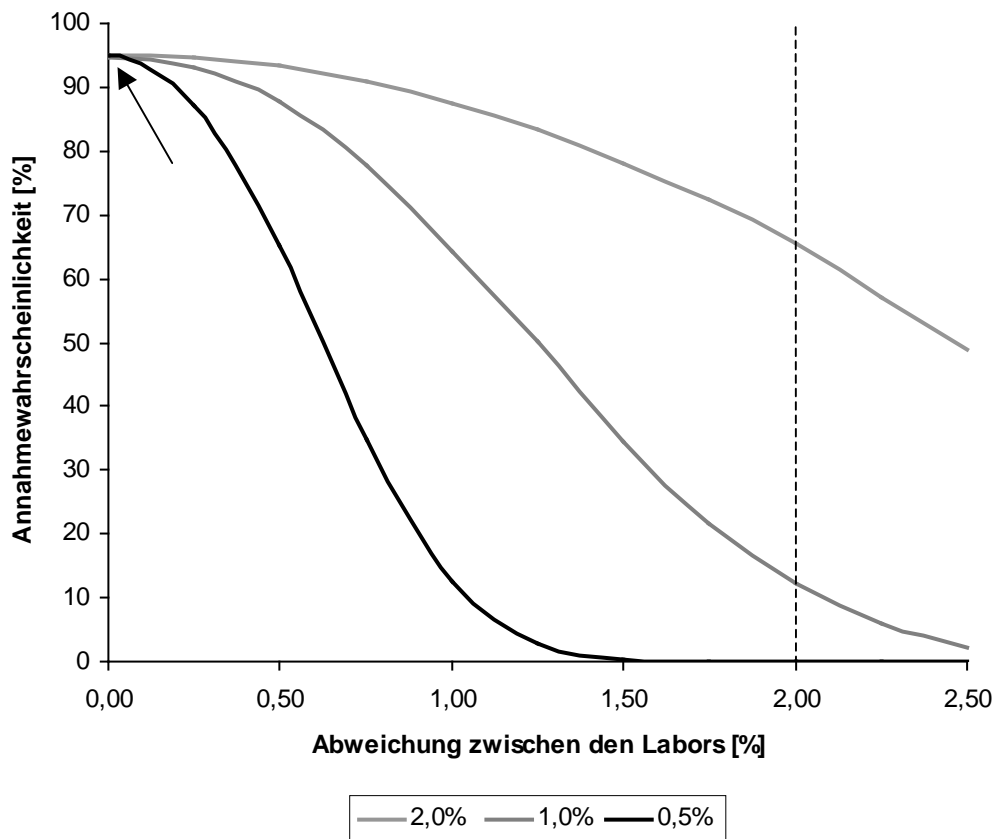


Abbildung 2.6-5: Annahmewahrscheinlichkeit des Zweistichproben-t-Testes in Abhängigkeit von der wahren Abweichung zwischen den Labors von 0 bis 2,5 %. Die Präzision als RSD% der drei Grundgesamtheiten betrug jeweils 0,5; 1,0 und 2,0 %; die Stichprobengröße $n = 6$.

Existiert keine Abweichung (s. Pfeil in Abb. 2.6-5), besitzen alle drei Transfersituationen, die sich nur in der Präzision voneinander unterscheiden, die gleiche Annahmewahrscheinlichkeit von 95 % und die gleiche Irrtumswahrscheinlichkeit von $\alpha = 5$ %. Bei einer wachsenden tatsächlichen Abweichung fallen die Annahmewahrscheinlichkeiten unterschiedlich stark ab.

Die Folgen aus der Abhängigkeit des Zweistichproben-t-Testes von der Präzision werden dann offenkundig, wenn man eine praktisch relevante Akzeptanzgrenze

($\theta_0 \pm \varepsilon$) wählt, welche eine maximal zulässige Abweichung kennzeichnet, die man nicht überschreiten möchte. Bei einer solchen Akzeptanzgrenze von z.B. 2 % zwischen den Labors (s. gestrichelte Linie in Abb. 2.6-5) bestehen dann sehr unterschiedliche Annahmewahrscheinlichkeiten, in Abhängigkeit der Präzision.

Im ersten Fall (0,5 % RSD%) liegt die Annahmewahrscheinlichkeit bei einer tatsächlichen Abweichung von 0,75 % bei nur 35 %, d.h. Transfersituationen, in denen sich die Labors zwar unterscheiden, dieser Unterschied aber noch tolerabel ist, werden mit hoher Wahrscheinlichkeit trotzdem abgelehnt.

Für eine RSD% von 1,0 % verläuft die Annahmewahrscheinlichkeit am günstigsten. Innerhalb des Akzeptanzbereichs ist die Annahmewahrscheinlichkeit hoch und darüber hinaus das Risiko einer fälschlichen Annahme nur mäßig erhöht.

Bei einer geringeren Präzision von 2,0 % hingegen kommt die oben erläuterte Problematik des breiten Konfidenzintervalls zum Tragen. Selbst an der Akzeptanzgrenze von 2 %, und vor allem darüber hinaus, werden die Transfers mit hoher Wahrscheinlichkeit (> 65 %) noch angenommen. Das Risiko, einen unzureichenden Transfer anzunehmen, ist in diesem Falle sehr hoch. Und gerade dieses schwerwiegendere Risiko kann dann durch den Zweistichproben-t-Test nicht kontrolliert werden.

Die Präzision, die den Zweistichproben-t-Test beeinflusst, ist meist nur näherungsweise bekannt und wird folglich geschätzt, so dass der tatsächliche Effekt auf die Leistungsfähigkeit des Zweistichproben-t-Testes unbekannt bleibt.

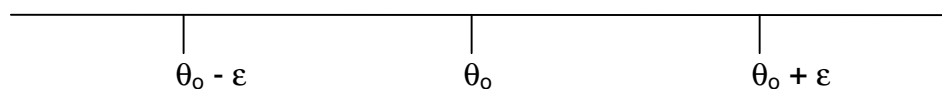
Um den Sollwert herum gibt es immer einen Bereich, innerhalb dessen ein Unterschied zwischen den Labors als praktisch irrelevant gilt. Paradoxerweise steigt die Wahrscheinlichkeit, anhand des Zweistichproben-t-Testes einen signifikanten Unterschied zwischen den Grundgesamtheiten zu erkennen, auch in diesem Bereich genau dann an, wenn sich die Präzision verbessert oder die Stichprobenzahl erhöht - und das, obwohl im Fall hoher Präzisionen oder großer Stichprobenzahlen ein höherer Informationsgehalt über die wahre Situation vorliegt. Das Konfidenzintervall wird aber in beiden Fällen schmaler und dadurch sinkt die Wahrscheinlichkeit, dass sich der Sollwert innerhalb des Intervalls befindet, unabhängig davon, wie klein die geschätzte Differenz ist (s. Abb. 2.6-3).

2.6.3 Konstruktionsprinzip von Äquivalenztests

Äquivalenztests unterscheiden sich vom klassischen t-Test und F-Test nicht durch die eingesetzten Teststatistiken, sondern durch die Form der Hypothesen (Äquivalenztest: Vergleich zweier Intervalle, t-Test: Vergleich von Intervall und Sollwert), deren Verifizierung anhand der auszuwertenden Daten angestrebt wird. Die allgemeine Grundidee für das Testen auf Äquivalenz wurde basierend auf einer Arbeit von Westlake [40] aus dem Jahre 1972 für Bioverfügbarkeitsstudien entwickelt [41]. Der Lösungsansatz für Äquivalenznachweisprobleme lässt sich am geeignetsten graphisch anhand von Konfidenzintervallen erklären:

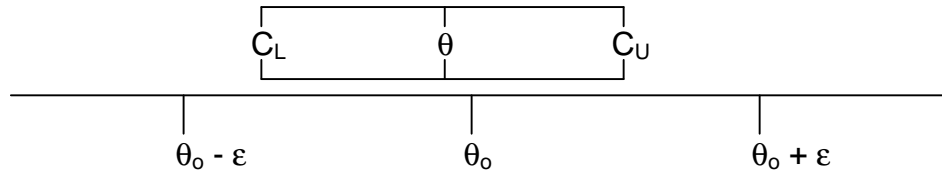
Für den jeweiligen zu testenden Hauptparameter θ ($\mu_1 - \mu_2$ oder $\hat{\sigma}_2^2 / \hat{\sigma}_1^2$) wird ein Konfidenzintervall gebildet. Die Äquivalenzhypothese, deren Plausibilität anhand der Stichproben beurteilt werden soll, behauptet die Gleichheit von θ mit einem geeigneten Sollwert θ_0 bis auf praktisch irrelevante Abweichungen. Dieser Sollwert θ_0 ist für das Testen der Differenz zweier Mittelwerte ($\mu_1 - \mu_2$) üblicherweise 0 und für das Testen des Quotienten zweier Varianzen ($\hat{\sigma}_2^2 / \hat{\sigma}_1^2$) die Zahl 1.

Zu diesem Sollwert θ_0 werden nun eine untere ($\theta_0 - \varepsilon$) und eine obere ($\theta_0 + \varepsilon$) Akzeptanzgrenze festgelegt, so dass ein symmetrisches Intervall $[\theta_0 - \varepsilon; \theta_0 + \varepsilon]$ um θ_0 entsteht:



Die Festlegung der Akzeptanzgrenzen (also von ε) kann durch Vorgaben seitens der Behörden oder auch durch firmeninterne Vorgaben bedingt sein. So ist z.B. beim Mittelwertsvergleich im Rahmen analytischer Methodentransfers von Methoden zur Gehaltsbestimmung ein ε von 2 % üblich, wodurch ein Intervall von $[-2\%; +2\%]$ resultiert [19].

Anhand der jeweiligen Teststatistik wird nun ein $(1-2\alpha)$ -Konfidenzintervall um den Parameter θ berechnet, welches von der unteren Grenze C_L („lower“) bzw. der oberen Grenze C_U („upper“) begrenzt wird. Die Größe dieses Konfidenzintervalls hängt von den Freiheitsgraden n und der Irrtumswahrscheinlichkeit α ab:



Das Testen der Hypothesen erfolgt über den Vergleich der Grenzen des Konfidenzintervalls mit den Akzeptanzgrenzen. Die Nullhypothese wird angenommen, wenn entweder $C_L \leq \theta_0 - \varepsilon$ oder $C_U \geq \theta_0 + \varepsilon$ gilt. In einem solchen Fall liegt das Konfidenzintervall zum Teil oder ganz außerhalb der Akzeptanzgrenzen (vgl. Abb. 2.6-6).

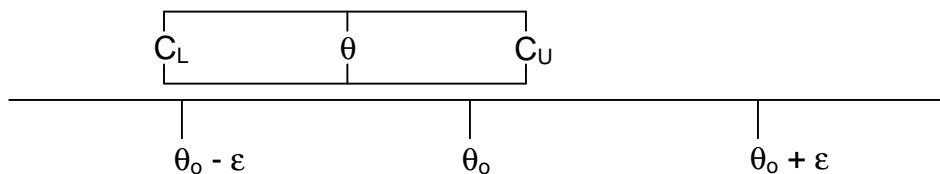


Abbildung 2.6-6: In dem hier abgebildetem Fall wird auf Nichtäquivalenz geschlossen, weil die untere Grenze des Konfidenzintervalls C_L kleiner als die untere Akzeptanzgrenze ($\theta_0 - \varepsilon$) ist ($C_L \leq \theta_0 - \varepsilon$).

Null- und Alternativhypothese lauten:

$$H_0: \theta \leq \theta_0 - \varepsilon \quad \vee \quad \theta \geq \theta_0 + \varepsilon$$

$$H_1: \theta_0 - \varepsilon < \theta < \theta_0 + \varepsilon.$$

Abgelehnt wird die Nullhypothese der Nichtäquivalenz genau dann, wenn das $(1-2\alpha)$ -Konfidenzintervall mit beiden Grenzen C_L bzw. C_U vollständig im hypothetischen Äquivalenzintervall $[\theta_0 - \varepsilon; \theta_0 + \varepsilon]$ enthalten ist, wenn also zugleich: $C_L > \theta_0 - \varepsilon$ und $C_U < \theta_0 + \varepsilon$ erfüllt sind [42]. In diesem Fall wird auf Äquivalenz geschlossen und die Alternativhypothese angenommen. Die geschätzte Lage des

Parameters samt Konfidenzintervall liegt signifikant innerhalb vorher gewählter, praktisch relevanter Grenzen (s. Abb. 2.6-7).

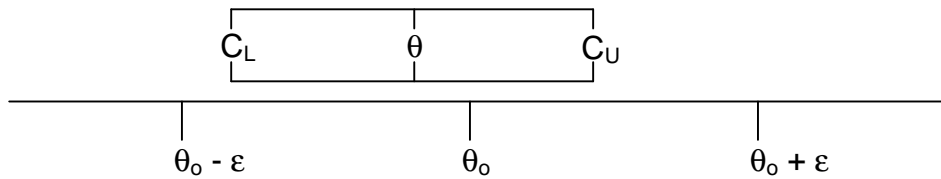


Abbildung 2.6-7: Das Schließen auf Äquivalenz erfolgt, wenn das durch C_U und C_L begrenzte Konfidenzintervall innerhalb der Grenzen $\theta_0 - \varepsilon$ und $\theta_0 + \varepsilon$ liegt.

Null- und Alternativhypothese sind also im Unterschied zum Zweistichproben-t-Test genau umgekehrt konstruiert. Während man beim Zweistichproben-t-Test durch Annahme der Nullhypothese auf Gleichheit schließt, so wird beim Äquivalenztest durch Annahme der Alternativhypothese auf Gleichheit geschlossen (sog. Hypothesenumkehr).

Ein weiterer Unterschied zum klassischen Zweistichproben-t-Test besteht darin, dass beim Äquivalenztest nicht getestet wird, ob der Sollwert θ_0 innerhalb des Konfidenzintervalls liegt oder nicht. Beim Äquivalenztest wird (graphisch gesehen) nur die Lage des Konfidenzintervalls bezüglich des durch die Akzeptanzgrenzen eingeschlossenen Intervalls betrachtet, welches beim t-Test keine Rolle spielt.

Zwar beträgt das zweiseitige *Konfidenzniveau* von C_L und C_U lediglich $(1 - 2\alpha)$. Dennoch hält der zugehörige Test immer das *Signifikanzniveau* α ein, (ein Beweis für diesen Sachverhalt wird in [42] gegeben). Die Wahrscheinlichkeit, einen nicht akzeptablen Unterschied anzunehmen, beträgt damit stets nicht mehr als α .

2.6.4 Äquivalenztest für den Mittelwertsvergleich

Das in Abschnitt 2.6.3 erläuterte allgemeine Prinzip wird im Folgenden für den Vergleich zweier Mittelwerte spezifiziert. Je nachdem, ob die Akzeptanzgrenzen, die man einsetzen möchte, in absoluter oder relativer Form vorliegen, sind unterschiedliche Formeln zu verwenden.

2.6.4.1 Einsatz absoluter Akzeptanzgrenzen

Die Testung zweier Mittelwerte auf Äquivalenz in den Simulationen im Rahmen dieser Arbeit (s. 7.1.3) erfolgte gegenüber absoluten Akzeptanzgrenzen. Der zugehörige Äquivalenztest läuft folgendermaßen ab [43] :

Die zu testenden Hypothesen lauten:

$$H_0: \mu_1 - \mu_2 \leq \theta_0 - \varepsilon \vee \mu_1 - \mu_2 \geq \theta_0 + \varepsilon \quad \text{gegen} \quad H_1: \theta_0 - \varepsilon < \mu_1 - \mu_2 < \theta_0 + \varepsilon.$$

Die zwei zu berechnenden Prüfgrößen lauten:

$$T_{t1} = \frac{(\bar{x}_1 - \bar{x}_2) - (\theta_0 - \varepsilon)}{\hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2.6-5)$$

$$T_{t2} = \frac{(\theta_0 + \varepsilon) - (\bar{x}_1 - \bar{x}_2)}{\hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2.6-6)$$

Diese Prüfgrößen unterscheiden sich vom klassischen t-Test (s. Gl. 2.6-1) dadurch, dass die Akzeptanzgrenze als Term $(\theta_0 \pm \varepsilon)$ mit in die Prüfgröße einbezogen wird. Als kritischer Wert wird $t_{n-1, n-1; 1-\alpha}$ verwendet mit α statt $\alpha/2$.

Die Nullhypothese wird verworfen, wenn beide Prüfgrößen $T_{t1,2} > t_{2n-2; 1-\alpha}$ sind und ansonsten angenommen. Aus Gl. 2.6-5 und 2.6-6 wird ersichtlich, dass sich der Äquivalenztest hier also als Kombination zweier einseitiger t-Tests auffassen lässt.

Zum besseren Vergleich wird das Beispiel aus 2.6.1 herangezogen:

$\bar{x}_1 = 99,67$; $\bar{x}_2 = 100,14$; $\hat{\sigma}_1^2 = 0,35$; $\hat{\sigma}_2^2 = 0,21$; $n_{1,2} = 6$; $\alpha = 0,05$. Zusätzlich sollen die absoluten Akzeptanzgrenzen bei $\theta_0 - \varepsilon = -2$ und $\theta_0 + \varepsilon = 2$ liegen.

$$T_{t1} = \frac{(99,67 - 100,14) + 2}{0,53 \sqrt{\frac{1}{6} + \frac{1}{6}}} = 5,00$$

$$T_{t2} = \frac{2 - (99,67 - 100,14)}{0,53 \sqrt{\frac{1}{6} + \frac{1}{6}}} = 8,07$$

Für den kritischen Wert ergibt sich: $t_{10;0,95} = 1,81$. T_{t1} und T_{t2} sind deutlich größer als 1,81. Daher wird die Nullhypothese verworfen und auf Äquivalenz geschlossen.

2.6.4.2 Einsatz relativer Akzeptanzgrenzen

Die Testung unter Verwendung relativer Akzeptanzgrenzen wie 2 % kann ebenfalls als Kombination zweier einseitiger t-Tests berechnet werden. Beim Testen der Hypothesen wird der Term $\theta_0 - \varepsilon$ in Form eines prozentualen Wertes mit in die Prüfgröße einbezogen (z.B. +2 % als 1,02 oder -2 % als 0,98).

Die zu testenden Hypothesen lauten:

$$H_0: \mu_1/\mu_2 \leq \theta_0 - \varepsilon \vee \mu_1/\mu_2 \geq \theta_0 + \varepsilon \quad \text{gegen} \quad H_1: \theta_0 - \varepsilon < \mu_1/\mu_2 < \theta_0 + \varepsilon.$$

Die Prüfgrößen unter Einbeziehung relativer Akzeptanzgrenzen sind durch folgende Gleichungen 2.6-7 und 2.6-8 gegeben [43]:

$$T_{t1} = \frac{(\bar{x}_2 - (\theta_0 - \varepsilon)\bar{x}_1)}{\hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{(\theta_0 - \varepsilon)^2}{n_2}}} \quad (2.6-7)$$

$$T_{t2} = \frac{(\theta_0 + \varepsilon)\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{(\theta_0 + \varepsilon)^2}{n_2}}} \quad (2.6-8)$$

Diese Prüfgrößen unterscheiden sich von Gl. 2.6-5 und 2.6-6 dadurch, dass der Term $(\theta_0 \pm \varepsilon)$ als relativer Wert mit in die Prüfgröße einbezogen wird. Als kritischer Wert wird ebenfalls $t_{n-1, n-1; 1-\alpha}$ verwendet.

Die Nullhypothese wird auch hier verworfen, wenn beide Prüfgrößen $T_{t1,2}$ größer als $t_{2n-2; 1-\alpha}$ sind und ansonsten angenommen.

Das Beispiel aus 2.6.1 und die relativen Akzeptanzgrenzen $\theta_0 - \varepsilon = 1 - 0,02 = 0,98$; $\theta_0 + \varepsilon = 1 + 0,02 = 1,02$ ergeben:

$$T_{t1} = \frac{(100,14 - 0,98 \cdot 99,67)}{0,53 \sqrt{\frac{1}{6} + \frac{0,98^2}{6}}} = 8,13 \quad T_{t2} = \frac{(1,02 \cdot 99,67) - 100,14}{0,53 \sqrt{\frac{1}{6} + \frac{1,02^2}{6}}} = 5,03$$

Da sowohl T_{t1} als auch T_{t2} größer als $t_{10;0,95} = 1,81$ sind, muss die Nullhypothese verworfen und auf Äquivalenz geschlossen werden.

2.6.4.3 Verwendung von relativen Akzeptanzgrenzen und eines Konfidenzintervalls

Die dritte Möglichkeit der Testung besteht in dem Vergleich von Prüfgröße und kritischem Wert gegen relative Akzeptanzgrenzen wie 2 % unter Verwendung eines Konfidenzintervalls. Graphisch gesehen wird untersucht, ob ein berechnetes 90%-Konfidenzintervall innerhalb der gewählten Akzeptanzgrenzen liegt (s. 2.6.3).

Kritischer Wert und die zu testenden Hypothesen lauten wie oben:

$$H_0: \mu_1/\mu_2 \leq \theta_0 - \varepsilon \vee \mu_1/\mu_2 \geq \theta_0 + \varepsilon \quad \text{gegen} \quad H_1: \theta_0 - \varepsilon < \mu_1/\mu_2 < \theta_0 + \varepsilon.$$

Die obere und untere Grenze (C_U , C_L) des Intervalls sind durch Gleichung 2.6-9 bis 2.6-11 gegeben [44]:

$$C_L = 100 \cdot [(\bar{x}_2 / \bar{x}_1) \cdot e^{-(t_{(2n-2, 1-\alpha)} \hat{\sigma})} - 1] \quad (2.6-9)$$

$$C_U = 100 \cdot [(\bar{x}_2 / \bar{x}_1) \cdot e^{(t_{(2n-2, 1-\alpha)} \hat{\sigma})} - 1] \quad (2.6-10)$$

$$\hat{\sigma} = \sqrt{\hat{\sigma}_p^2 \cdot \left(\frac{1}{n_1 \bar{x}_1^2} + \frac{1}{n_2 \bar{x}_2^2} \right)} \quad (2.6-11)$$

Die hier berechneten Größen werden (anders als in Gl. 2.6-7 und 2.6-8) nicht mit dem kritischen Wert verglichen, sondern direkt mit den prozentualen

Akzeptanzgrenzen. Der kritische Wert ist in den Gleichungen 2.6-9 und 2.6-10 bereits enthalten. Zur Berechnung von $\hat{\sigma}_p^2$ siehe Gl. 2.6-2. Die Nullhypothese wird angenommen, wenn zugleich: $C_L > \theta_0 - \varepsilon$ und $C_U < \theta_0 + \varepsilon$ sind.

Das Beispiel aus 2.6.1:

$\bar{x}_1 = 99,67$; $\bar{x}_2 = 100,14$; $\hat{\sigma}_1^2 = 0,35$; $\hat{\sigma}_2^2 = 0,21$; $n_1 = n_2 = 6$; $\theta_0 - \varepsilon = -2$; $\theta_0 + \varepsilon = 2$; $\alpha = 0,05$ ergibt:

$$C_L = 100 \cdot [(100,14/99,67) \cdot e^{-(1,81 \cdot 0,003)} - 1] = -1,02$$

$$C_U = 100 \cdot [(100,14/99,67) \cdot e^{(1,81 \cdot 0,003)} - 1] = 0,08$$

$$\hat{\sigma} = \sqrt{0,53^2 \cdot \left(\frac{1}{6 \cdot 99,67^2} + \frac{1}{6 \cdot 100,14^2} \right)} = 0,003$$

Da die untere Intervallgrenze C_L mit -1,02 und die obere Grenze C_U mit 0,08 beide im geforderten Intervall von $[-2; 2]$ liegen, wird die Nullhypothese verworfen und auf Äquivalenz geschlossen.

Die drei Berechnungsweisen unterscheiden sich in dem Einsatz der Art der Akzeptanzgrenzen. Die Berechnung anhand des selben Beispiels zeigt, dass sie zu numerisch unterschiedlichen Rechenergebnissen führen. Trotzdem liefern sie das gleiche logische Ergebnis im Hinblick auf die Annahme oder Ablehnung der Nullhypothese. Der Vorteil der dritten Methode aus 2.6.4.3 liegt in der Zusatzinformation des bei der Berechnung gebildeten Konfidenzintervalls. Das Konfidenzintervall lässt erkennen, ob die Akzeptanzgrenzen deutlich oder nur knapp eingehalten werden. Dies ist bei der Berechnung der Prüfwerte in 2.6.4.1 und 2.6.4.2 schwieriger zu erkennen.

2.6.5 Vorteil des Äquivalenztestes gegenüber dem Zweistichproben-t-Test

Die analytische Fragestellung im Rahmen analytischer Methodentransfers sollte lauten: „Unterscheiden sich die Mittelwerte um mehr als eine irrelevante Differenz ε voneinander?“ Der Äquivalenztest wird dieser Fragestellung gerecht durch die Formulierung der Nullhypothese $H_0: \mu_1 - \mu_2 \leq \theta_0 - \varepsilon \vee \mu_1 - \mu_2 \geq \theta_0 + \varepsilon$ (s. Abs. 2.6.3). Wird diese Nullhypothese fälschlicherweise abgelehnt, so tritt auch hier ein α -Fehler ein. In diesem Fall wird ein eigentlich unzureichender Transfer zu Unrecht angenommen, ein Risiko, das auch als Konsumentenrisiko bezeichnet wird. Dieses bedeutendere Risiko, eine unzureichende Analytik in der Routinefreigabe zu übersehen, wird durch den Äquivalenztest unabhängig von Präzision und Stichprobengröße auf einen transparenten Wert von z.B. 5 % gehalten (s. Abb. 2.6.8).

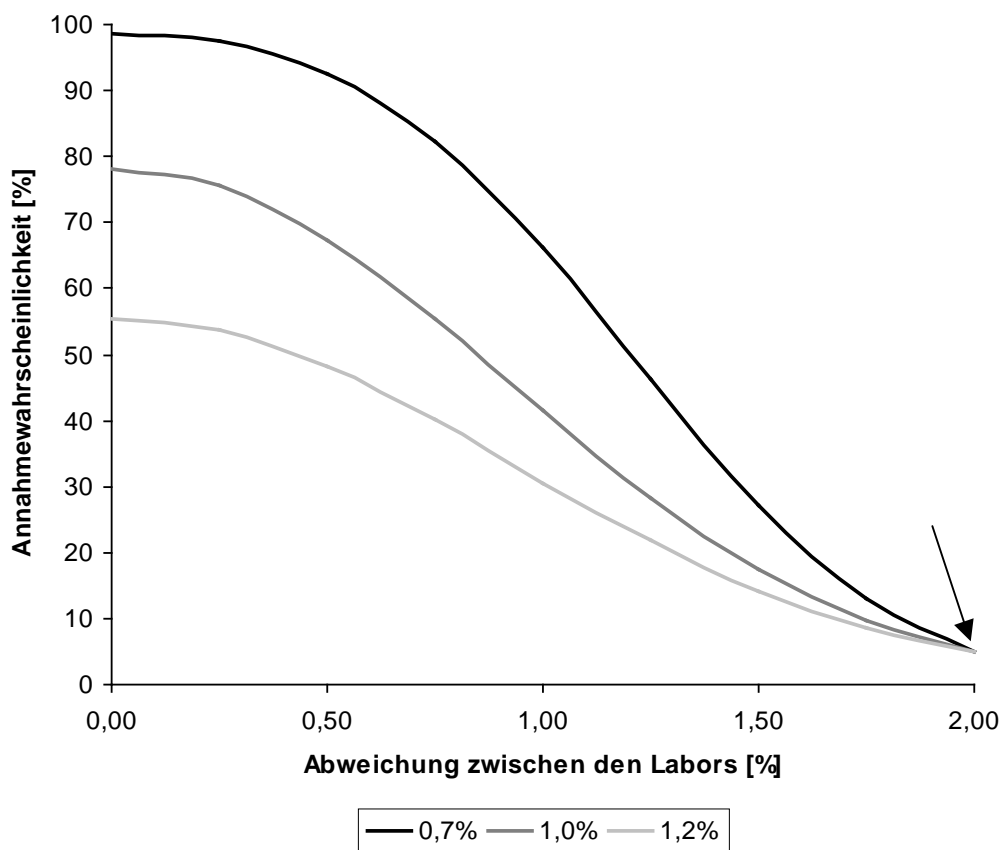


Abbildung 2.6-8: Annahmewahrscheinlichkeit des Äquivalenztestes mit einer Akzeptanzgrenze von 2 % in Abhängigkeit der wahren Abweichung zwischen den Labors von 0 bis 2 %. Die drei Transfersituationen unterscheiden sich nur in der Präzision als RSD% der beteiligten Labors (0,7; 1,0 und 1,2 %).

Der Unterschied des Äquivalenztestes zum Zweistichproben-t-Test ist in Abbildung 2.6.8 deutlich zu sehen. An der Akzeptanzgrenze von 2 % (s. Pfeil in Abb. 2.6.8) wird in allen drei Fällen die Annahmewahrscheinlichkeit unabhängig von der Präzision auf 5 % kontrolliert. Das untergeordnetere Risiko des β -Fehlers, einen eigentlich erfolgreichen Transfer zu übersehen und einen Methodentransfer erneut durchführen zu müssen, hängt von der Stichprobengröße und der Präzision der beteiligten Labors ab.

Die Annahmewahrscheinlichkeit $1-\beta$ (Power) und der β -Fehler ergeben sich innerhalb der Akzeptanzgrenzen zu 100 %. Steigt die Power an, so sinkt entsprechend der β -Fehler.

Der Einfluss der Präzision auf die Power unterscheidet sich beim Äquivalenztest vom Zweistichproben-t-Test: Beim Äquivalenztest wird die Power größer, wenn sich die Präzision verbessert. So beträgt für eine wahre Abweichung von 0,50 % und einer Laborpräzision von 1,0 % die Power 68 %. Für eine Laborpräzision von 0,7 % steigt die Power bei gleicher Abweichung auf 93 % an (s. Abb. 2.6.8).

Liegt - bedingt durch eine hohe Stichprobenzahl oder eine bessere Analytik - eine hohe Präzision vor, dann steigt die Wahrscheinlichkeit, einen Transfer zurecht anzunehmen. Das ist ein im Gegensatz zum Zweistichproben-t-Test nachvollziehbarer Zusammenhang, denn eine hohe Präzision bedeutet schließlich einen höheren statistischen Informationsgehalt.

Abbildung 2.6.9 zeigt in einer Übersicht noch einmal die prinzipiell unterschiedliche Leistung von Äquivalenztest und Zweistichproben-t-Test in Abhängigkeit von der wahren Abweichung zwischen den Labors.

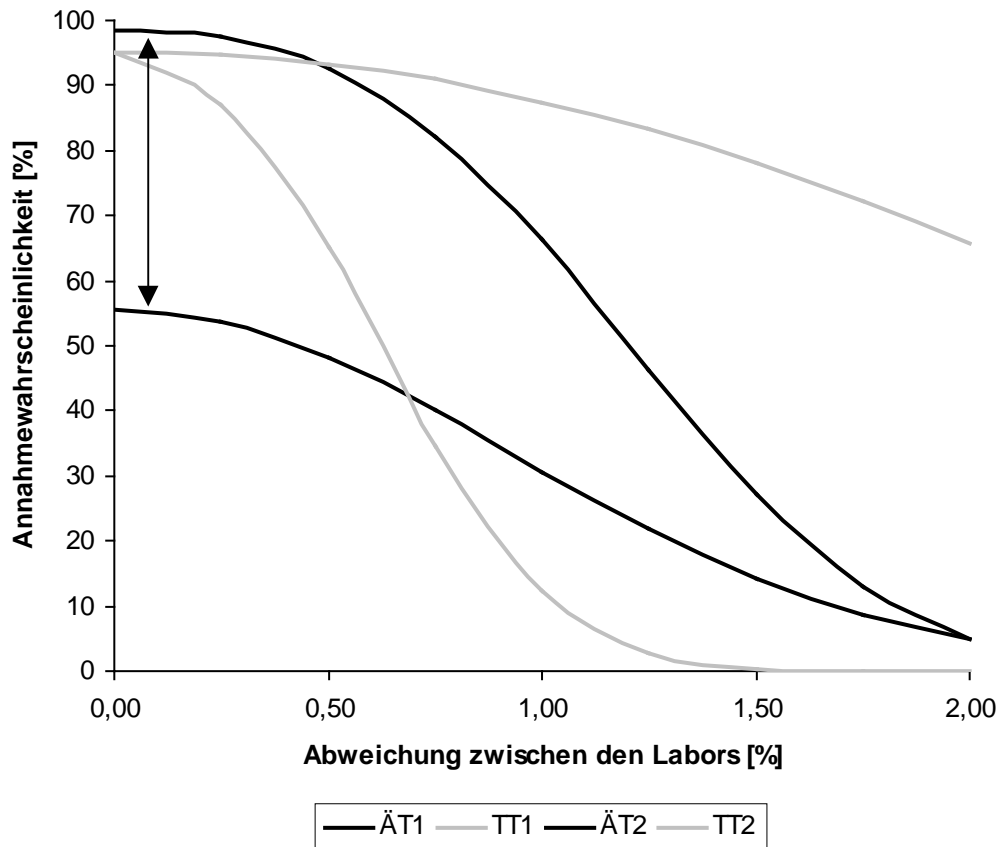


Abbildung 2.6-9: Übersicht über die grundsätzlich unterschiedlichen Leistungen von Zweistichproben-t-Test **TT** und Äquivalenztest **ÄT** (ÄT1 und ÄT2 mit $n = 5$ und 0,7 bzw. 1,2 % RSD%; TT1 und TT2 mit $n = 6$ und 0,5 bzw. 2,0 % RSD%).

Der Zweistichproben-t-Test kontrolliert das Risiko, einen Transfer fälschlicherweise zu übersehen, an der Stelle Bias = 0 auf 5 %. Der Begriff Bias steht für die Differenz zwischen dem Schätzer und dem Parameter θ (z.B. $\bar{x} - \mu$). Der Äquivalenztest hingegen kontrolliert das bedeutendere Risiko, einen Transfer fälschlicherweise anzunehmen, auf 5 %, sobald die wahre Abweichung die Akzeptanzgrenze überschritten hat. Der durch den Doppelpfeil markierte Bereich zeigt an, dass beim Äquivalenztest die Leistungsfähigkeit (als Power) je nach Transfersituation unterschiedlich sein kann. Daher bestand ein Ziel dieser Arbeit

unter anderem darin, für gängige Fehlerbeträge und Versuchsdesigns diese Leistungsfähigkeit anhand von Simulationen zu ermitteln und darzustellen.

2.6.6 Nichtstatistischer Vergleich der Mittelwerte

Als Alternative zu den statistischen Testungen wird der nichtstatistische Vergleich der beiden Mittelwerte zur Datenauswertung vorgeschlagen [21]. Diese sehr einfache Testung prüft, ob die Differenz der Mittelwerte der beiden beteiligten Labors innerhalb vorher gewählter Akzeptanzgrenzen liegt:

Die zu testenden Hypothesen lauten:

$$H_0: -2\% < \bar{x}_R - \bar{x}_T < 2\% \quad \text{gegen} \quad H_1: \bar{x}_R - \bar{x}_T < -2\% \vee \bar{x}_R - \bar{x}_T > 2\%$$

Liegt der Term $\bar{x}_R - \bar{x}_T$ innerhalb von $\pm 2\%$, so wird der Transfer angenommen und ansonsten abgelehnt.

Die Leistungsfähigkeit lässt sich ebenfalls durch Simulationen bestimmen und ist in Abbildung 2.6-10 exemplarisch dargestellt:

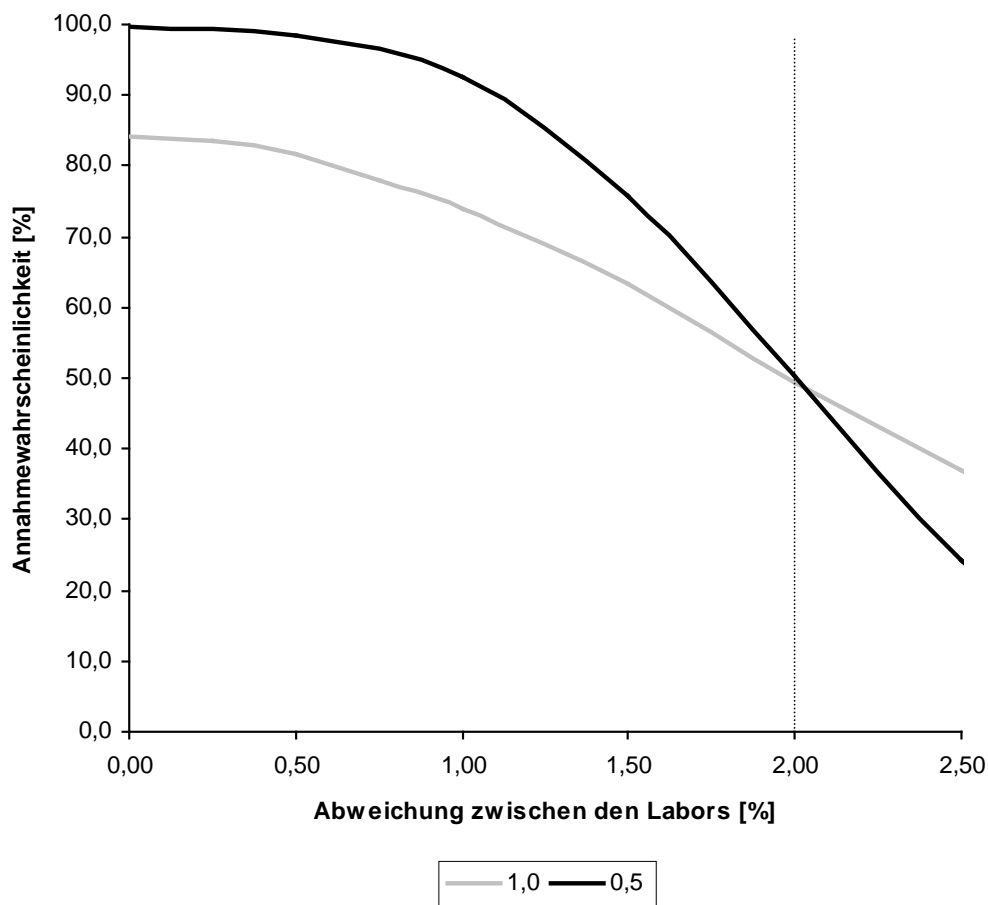


Abbildung 2.6-10: Annahmewahrscheinlichkeit zweier Transfersituationen mit unterschiedlichen Präzisionen der beteiligten Labors (1,0 % und 0,5 % RSD%).

Als Akzeptanzgrenze dient in diesem Beispiel das Intervall von $\pm 2\%$. Die beiden Transfersituationen unterscheiden sich nur in der Präzision der beteiligten Labors voneinander. Die Annahmewahrscheinlichkeit an der Akzeptanzgrenze bei einem tatsächlichen Unterschied von 2% (gestrichelte Linie) beträgt in beiden Fällen 50% und ist damit unabhängig von der Präzision.

Die Transfersituation mit der höheren Präzision ($0,5\%$) besitzt innerhalb des Akzeptanzintervalls vergleichsweise höhere Annahmewahrscheinlichkeiten, d.h. je höher die Präzision, umso leistungsfähiger der Mittelwertvergleich.

Eine Annahmewahrscheinlichkeit von 50% direkt an der Akzeptanzgrenze ist nicht akzeptabel, aber durch geschickte Wahl einer geringeren Akzeptanzgrenze als eigentlich gewünscht, ist ein günstigerer Verlauf der Annahmewahrscheinlichkeit zu erreichen.

So führt z.B. bei der Transfersituation, in der beide Labors eine Präzision von $0,37\%$ besitzen, die Wahl einer tatsächlichen Akzeptanzgrenze von $1,10\%$ dazu, dass an der Stelle 2% eine sinnvolle Annahmewahrscheinlichkeit von ca. 5% existiert, wie in Abbildung 2.6-11 gezeigt.

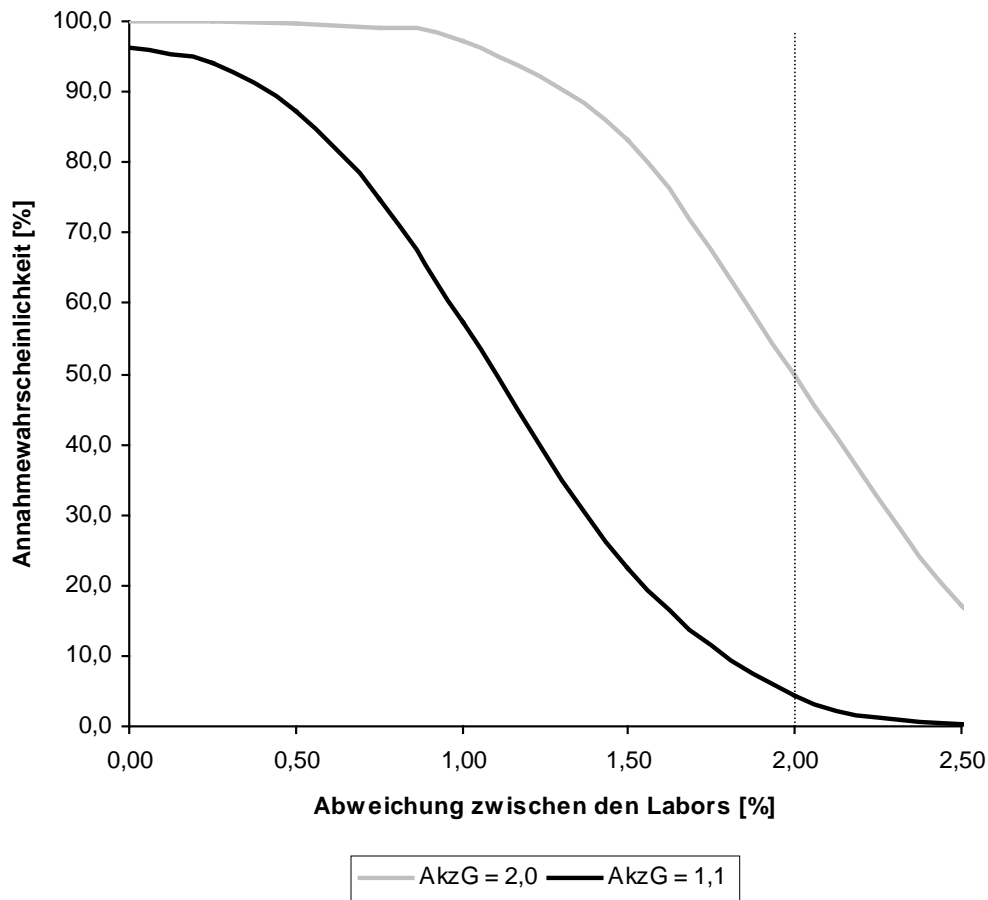


Abbildung 2.6-11: Annahmewahrscheinlichkeit zweier Situationen mit unterschiedlichen Akzeptanzgrenzen.

Es ist zu sehen, dass die Wahl einer kleineren Grenze als benötigt zu sinnvollen Annahmewahrscheinlichkeiten führen kann. Im Vergleich zum Äquivalenztest ist aber auch durch geschickte Wahl der Grenzen beim Mittelwertvergleich keine wesentliche Verbesserung der Wahrscheinlichkeiten zu erzielen, so dass der Nachteil der nur indirekten Kontrolle der fälschlichen Annahme eines Transfers hier überwiegt. Eine Unterschätzung der Präzision führt zu ungewollt hohen Annahmewahrscheinlichkeiten an der Akzeptanzgrenze. Daher ist dem Äquivalenztest auch gegenüber dem Mittelwertvergleich der Vorzug zu geben.

2.6.7 Klassischer F-Test

Der statistische Vergleich der Varianzen zweier unabhängiger Messreihen lässt sich anhand des F-Testes vornehmen [25, 32, 45]. Dabei wird geprüft, ob zwischen den Varianzen der zugrundeliegenden Grundgesamtheiten ein statistisch signifikanter Unterschied existiert.

Im Rahmen eines analytischen Methodentransfers ist nur die einseitige Fragestellung interessant, ob die geschätzte Varianz des Ziellabors größer als die des Referenzlabors ist, oder nicht. Denn eine bessere Präzision des Ziellabors ist als unproblematisch anzusehen.

Die Hypothesen lauten folgendermaßen:

$H_0: \hat{\sigma}_1^2 = \hat{\sigma}_2^2$ gegen $H_1: \hat{\sigma}_1^2 \neq \hat{\sigma}_2^2$. $\hat{\sigma}_1^2$ stellt die geschätzte Varianz des Referenz- und $\hat{\sigma}_2^2$ die des Ziellabors dar, wobei $\hat{\sigma}_2^2 > \hat{\sigma}_1^2$ sein muss. Ist $\hat{\sigma}_2^2$ kleiner als $\hat{\sigma}_1^2$, kann auf die Durchführung des F-Testes, wie oben erwähnt, verzichtet werden.

Die Prüfgröße T_F berechnet sich nach Gleichung 2.6-12:

$$T_F = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \quad (2.6-12)$$

Die Prüfgröße T_F wird mit dem kritischen Wert $F_{n_2-1, n_1-1; 1-\alpha}$ verglichen. Ist die $T_F > F_{n_2-1, n_1-1; 1-\alpha}$, dann wird die H_0 zum Niveau α verworfen und auf einen signifikanten Unterschied geschlossen. Anderenfalls wird H_0 angenommen. Die Leistungsfähigkeit des Testes ist für eine Stichprobengröße von $n = 20$ und einer Irrtumswahrscheinlichkeit $\alpha = 0,10$ exemplarisch in Abbildung 2.6-12 dargestellt.

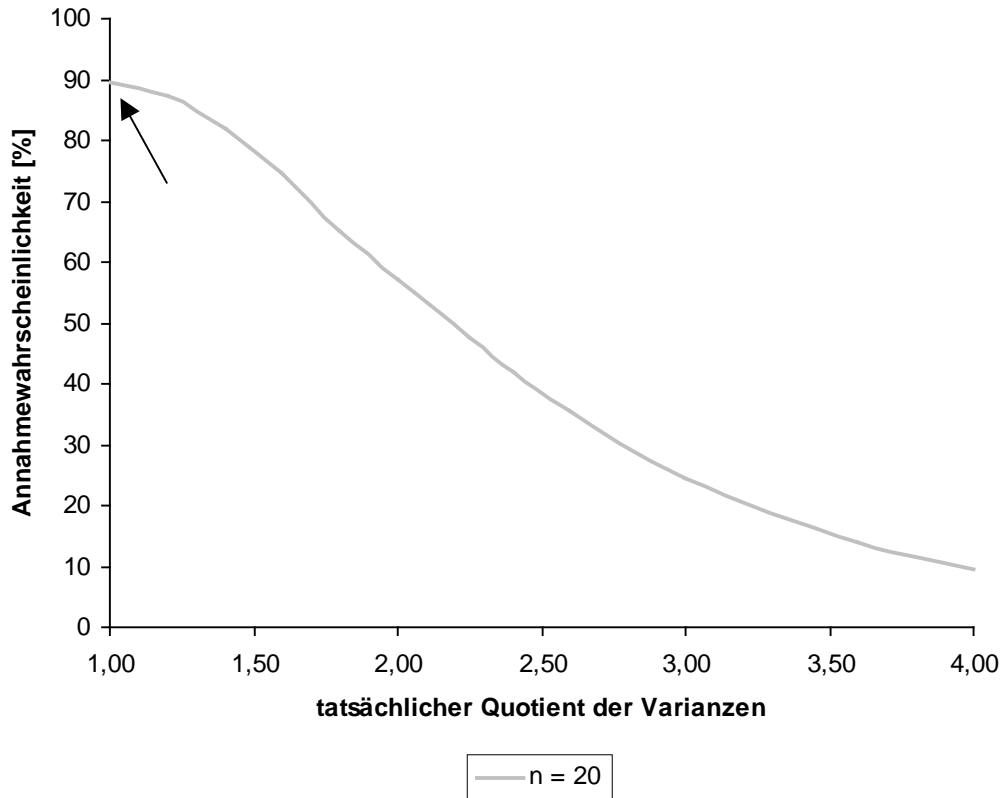


Abbildung 2.6-12: Annahmewahrscheinlichkeit des F-Testes gegen den Quotienten zweier Varianzen ($\alpha = 10\%$).

Zu sehen ist das Abfallen der Annahmewahrscheinlichkeit des F-Testes ausgehend von gleichen Varianzen ($\hat{\sigma}_2^2 / \hat{\sigma}_1^2 = 1$) bis hin zum Vierfachen ($\hat{\sigma}_2^2 / \hat{\sigma}_1^2 = 4$). Besteht in Wahrheit keine Abweichung zwischen den Labors (s. Pfeil), so beträgt die Wahrscheinlichkeit, die Alternativhypothese anzunehmen und sich dabei zu irren, genau $\alpha = 10\%$ (gewählte Irrtumswahrscheinlichkeit).

2.6.8 Nachteil des F-Testes

Der F-Test kann den β -Fehler, d.h. das Risiko, einen nicht akzeptablen Unterschied zwischen den Varianzen beider Labors zu übersehen, nicht präzisionsunabhängig kontrollieren. Dieser Nachteil des F-Testes ist in Abbildung 2.6-13 dargestellt.

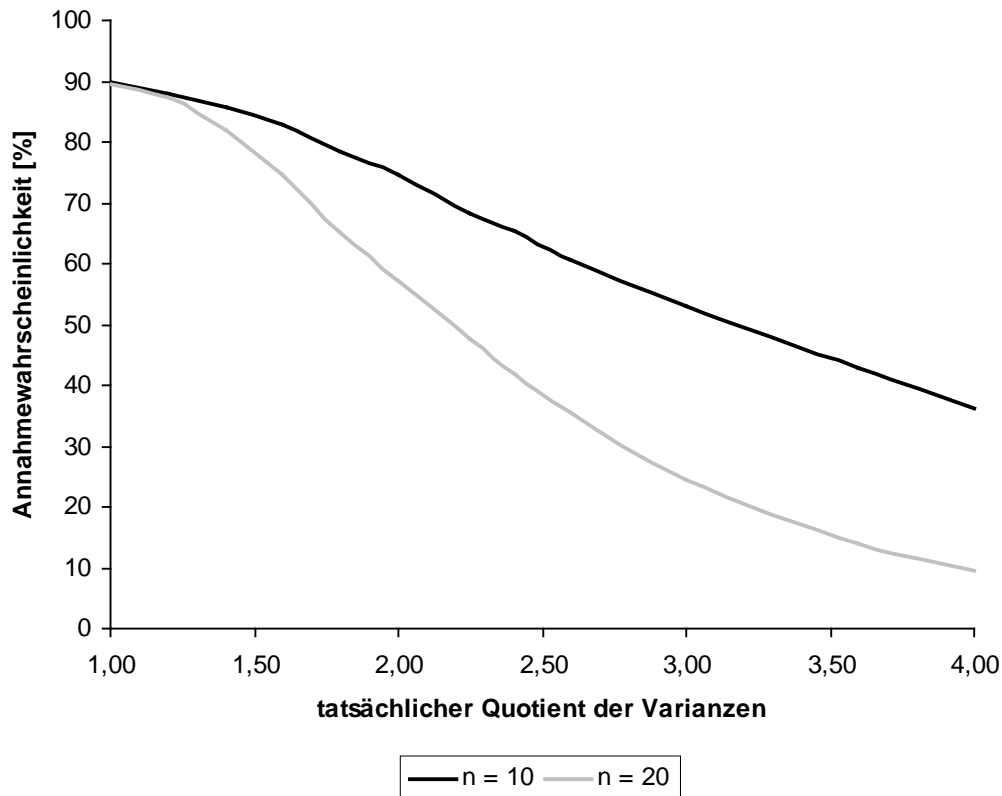


Abbildung 2.6-13: Verlauf der Power des F-Testes bei unterschiedlichem n . Die Situation ist die gleiche wie in Abb. 2.6-11. Paradoxerweise sinkt die Power, wenn n größer wird.

Die praktisch relevante Akzeptanzgrenze in Abb. 2.6-13 soll in diesem Beispiel bei $\hat{\sigma}_2^2 / \hat{\sigma}_1^2 = 4$ liegen, so dass die Annahmewahrscheinlichkeit des ganzen Bereichs die Power darstellt. Es ist zu sehen, dass an der Akzeptanzgrenze (hier $\hat{\sigma}_2^2 / \hat{\sigma}_1^2 = 4$) die Annahmewahrscheinlichkeit durch die Stichprobengröße nur indirekt kontrolliert wird. Bei einer Stichprobengröße von $n = 10$ beträgt die Power ca. 40 % bei $n = 20$ ca. 10 %.

Gleichzeitig sinkt die Power mit zunehmender Stichprobengröße, d.h. die Wahrscheinlichkeit, den F-Test zurecht zu akzeptieren, wird umso geringer, je größer der Stichprobenumfang ist.

An der Stelle $\hat{\sigma}_2^2 / \hat{\sigma}_1^2 = 2$ und bei einer Stichprobengröße von $n = 10$ beträgt die Power 75 %. Für $n = 20$ hingegen sinkt die Power auf 57 % ab. Obwohl im zweiten Fall mehr Information vorliegt, wird häufiger zu Unrecht abgelehnt. Bei hohen

Stichprobenzahlen wird dann zwar das Risiko eines β -Fehlers geringer, aber gleichzeitig werden zu oft erfolgreiche Transfers abgelehnt.

2.6.9 Äquivalenztest für den Vergleich zweier Varianzen

Der Äquivalenztest für den Vergleich zweier unabhängiger Varianzen wird prinzipiell so durchgeführt, dass das zweiseitige 90%-Konfidenzintervall (s. 2.6.3) des Quotienten beider Varianzen gebildet und mit im Vorfeld für praktisch relevant gehaltenen Grenzen verglichen wird. Um auf Äquivalenz zu schließen, muss das Intervall innerhalb dieser Grenzen liegen [46].

Da, wie bereits erwähnt, im Rahmen von analytischen Methodentransfers nur eine im Vergleich zu geringe Präzision unerwünscht ist, wird nur die obere Akzeptanzgrenze des Konfidenzintervalls gegen die Grenze A getestet.

Geprüft wird auf einen Unterschied der den beiden Stichproben zugrunde liegenden Varianzen der Grundgesamtheiten:

$$H_0: \hat{\sigma}_1^2 \neq \hat{\sigma}_2^2 \quad \text{gegen} \quad H_1: \hat{\sigma}_1^2 = \hat{\sigma}_2^2.$$

Die Prüfgröße T_F berechnet sich nach Gleichung 2.6-13:

$$T_F = \frac{\hat{\sigma}_2^2 / \hat{\sigma}_1^2}{F_{n_2-1, n_1-1; \alpha}} \leq A \quad (2.6-13)$$

Hierbei stellen wiederum $\hat{\sigma}_1^2$ die Varianz des Referenzlabors und $\hat{\sigma}_2^2$ die des Ziellabors dar. Die obere Akzeptanzgrenze A stellt denjenigen Unterschied zwischen den Varianzen dar, der als gerade noch vertretbar angesehen wird.

$F_{n_2-1, n_1-1; \alpha}$ ist das Quantil der F-Verteilung mit $(n_2 - 1)$ und $(n_1 - 1)$ Freiheitsgraden.

Üblicherweise wird auch hier $\alpha = 0,05$ gewählt.

Die Prüfgröße T_F wird mit der Akzeptanzgrenze A verglichen. Ist die $T_F < A$, dann wird die H_0 zum Niveau α verworfen und auf Äquivalenz geschlossen und ansonsten angenommen. Als praktisch relevanter Unterschied zwischen den Varianzen wird z.B. der Faktor 4,0 vorgeschlagen [46]. Auf der Ebene von

Standardabweichungen entspricht das dem Fall, dass die doppelte Standardabweichung des Ziellabors im Vergleich zum Referenzlabor noch gerade zulässig wäre.

Beispiel: Eine Situation mit $\hat{\sigma}_1^2 = 2,84$, $\hat{\sigma}_2^2 = 3,87$, $A = 4$, $\alpha = 0,05$ und $n_{1,2} = 12$ ergibt:

$$T_F = \frac{3,87^2 / 2,84^2}{F_{11,11;0,05}} = 3,89$$

Da die Prüfgröße T_F mit 3,89 ($F_{11,11;0,05} = 4,0$) etwas kleiner als 4,0 ist, wird die Alternativhypothese angenommen und auf Äquivalenz geschlossen.

2.6.10 Vorteil des Äquivalenztestes im Vergleich zum F-Test

Wie beim Vergleich zweier Mittelwerte kontrolliert der Äquivalenztest auch für den Vergleich zweier unabhängiger Varianzen den α -Fehler auf das gewählte Niveau 5 %. Abbildung 2.6-14 zeigt ein Beispiel für den Verlauf zweier Transfersituationen bei unterschiedlichen Stichprobenzahlen.

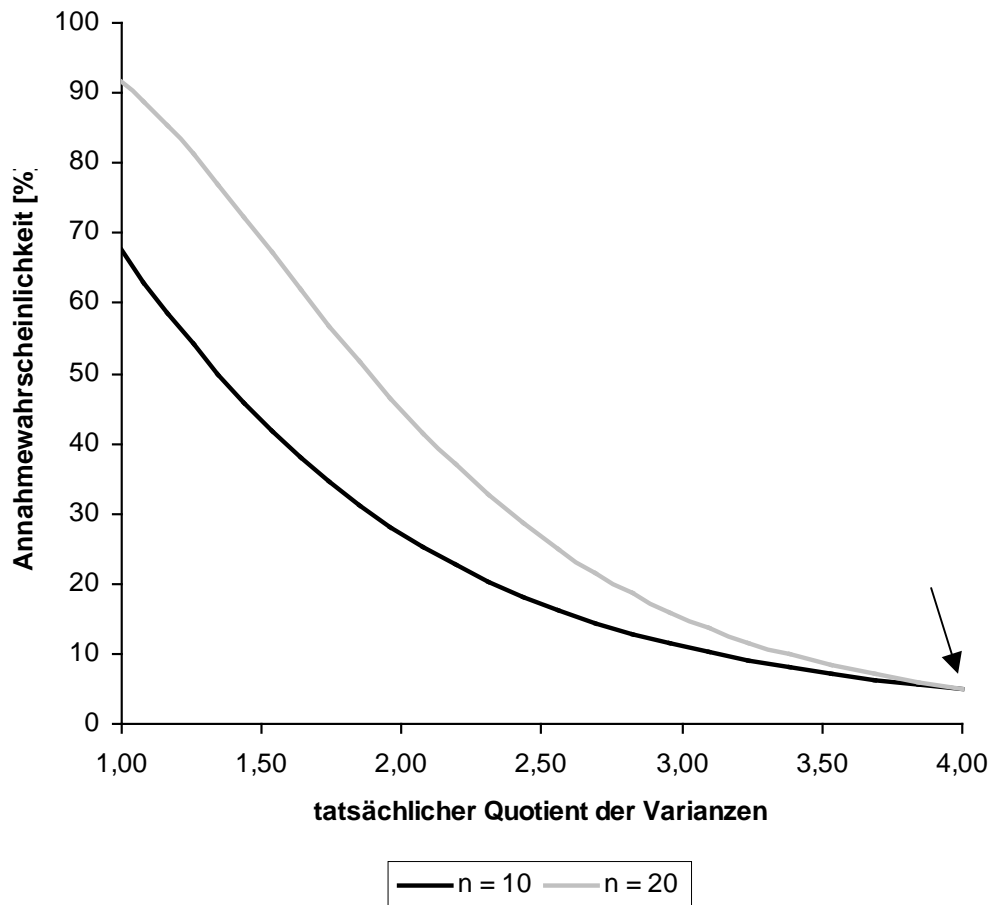


Abbildung 2.6-14: Verlauf der Annahmewahrscheinlichkeit des Äquivalenztestes für den Vergleich zweier Varianzen bei unterschiedlichen Stichprobenzahlen n ($\alpha = 0,05$; Akzeptanzgrenze = 4,00). Die Annahmewahrscheinlichkeit steigt bei zunehmendem n an. An der Akzeptanzgrenze wird das Risiko auf 5 % kontrolliert.

Die Power ($1-\beta$) im Annahmebereich steigt mit der Stichprobengröße. In welchem Maße die Power als Kriterium der Leistungsfähigkeit des Äquivalenztestes von der Stichprobengröße und der Akzeptanzgrenze abhängt, wird in Kapitel 4 eingehend diskutiert.

Abbildung 2.6-15 macht in einer Übersicht die prinzipiell unterschiedliche Leistung von Äquivalenztest und F-Test in Abhängigkeit von der wahren Abweichung zwischen den Labors deutlich.

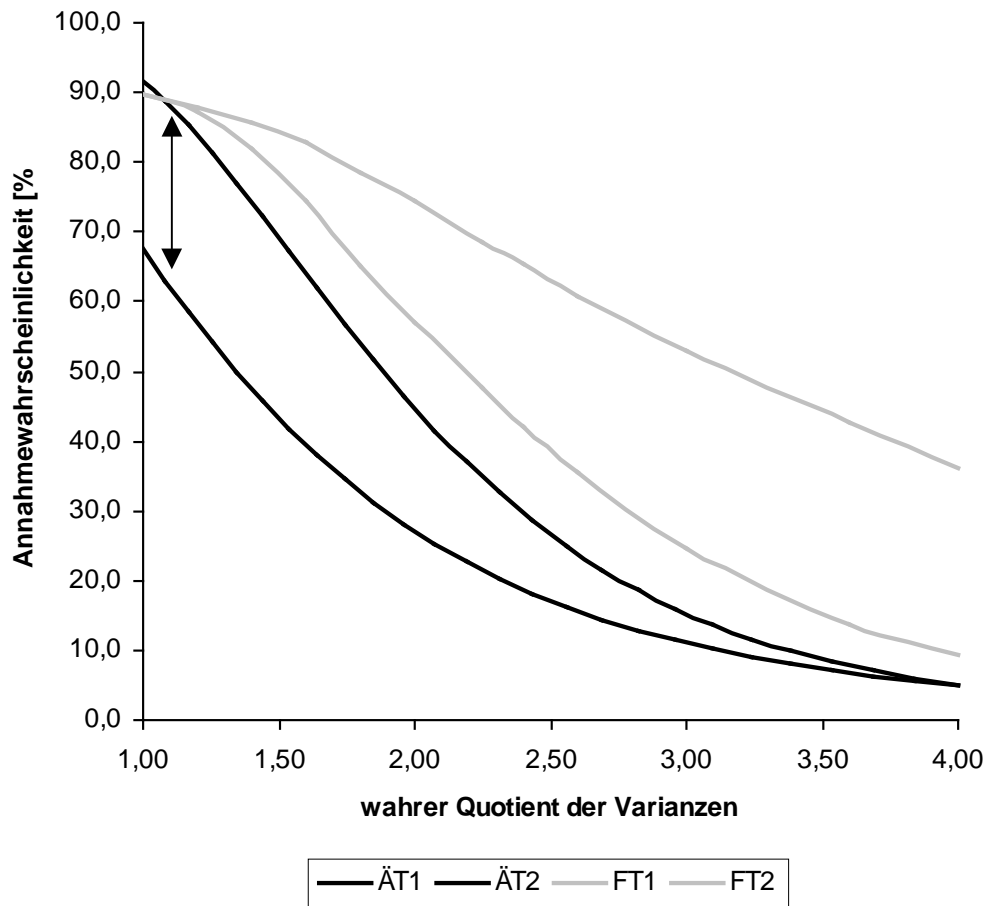


Abbildung 2.6-15: Übersicht über die unterschiedlichen Leistungen vom F-Test (FT) und Äquivalenztest (ÄT). Der F-Test kontrolliert das Risiko, einen Transfer fälschlicherweise zu übersehen, an der Stelle Bias = 0 (in diesem Falle auf 10 %). Der Äquivalenztest hingegen kontrolliert das bedeutendere Risiko, einen Transfer fälschlicherweise anzunehmen, an der Akzeptanzgrenze 4,00 (hier auf 5 %).

2.6.11 Testung der Standardabweichung gegen eine Akzeptanzgrenze

Ist im Rahmen eines analytischen Methodentransfers eine Obergrenze für die Präzision der zu transferierenden Methode bekannt, kann als weitere Alternative zu dem auf dem F-Test basierenden Äquivalenztest die Präzision als Standardabweichung (oder zur einfacheren Vergleichbarkeit als prozentuale relative Standardabweichung) gegen eine vorher festgelegte Akzeptanzgrenze getestet werden [47].

Die Nullhypothese $H_0: \hat{\sigma}^2 \geq A$ wird angenommen, wenn der Prüfterm größer als die Akzeptanzgrenze ist und anderenfalls zugunsten der Alternativhypothese $H_1: \hat{\sigma}^2 < A$ abgelehnt. Im Rahmen dieses Testes wird die obere Grenze des um die geschätzte Standardabweichung gebildeten Konfidenzintervalls berechnet und mit einer Akzeptanzgrenze A nach Gleichung 2.6-14 verglichen:

$$\hat{\sigma} \sqrt{\frac{n-1}{\chi^2_{\alpha, n-1}}} < A \quad (2.6-14)$$

Hierbei wird analog zum Äquivalenztest für den Vergleich zweier Varianzen die ungewünschte Annahme einer zu hohen Standardabweichung auf 5 % kontrolliert. Die Leistungsfähigkeit ist für gängige Fehlerbeträge anhand von Simulationen ermittelt worden und wird in Abschnitt 4.4 dargestellt.

3 Präzision und Wiederfindung in der HPLC

3.1 Methodenoptimierung

Die in den Messreihen eingesetzte Methode (s. 6.4) ist ausgehend von einer bereits existierenden Methode weiterentwickelt worden. Gegenüber der ursprünglichen Methode sollten zwei Dinge optimiert werden: Die mittlere Retentionszeit von Glibenclamid war mit mehr als 20 min relativ lang und die Stabilität des Glibenclamids in der Probenlösung mit drei Stunden bei 22 °C nicht ausreichend für längere Messserien. Daher galt es, unter Beibehaltung der Trennung des Glibenclamids (1) von den Nebenkomponten (2-6) (s. Abb. 3.2-8) die Retentionszeit zu verkürzen und die Probenstabilität zu erhöhen. Abbildung 3.1-1 zeigt ein Chromatogramm nach Anwendung der Standardmethode.

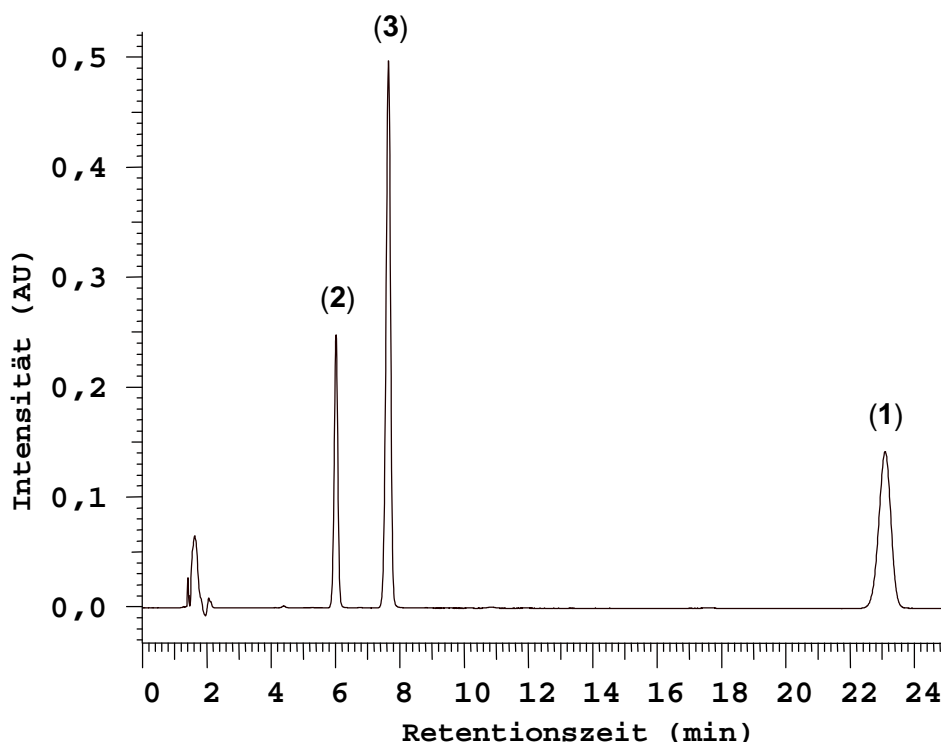


Abbildung 3.1-1: Chromatogramm der Ausgangsmethode. Glibenclamid (1) ist deutlich von den Komponenten (2) und (3) getrennt. Die lange Retentionszeit des Glibenclamids und die Tatsache, dass sich Glibenclamid mit der Zeit zu (2) zersetzt, machten mehrere Entwicklungsschritte notwendig.

3.1.1 Verkürzung der Retentionszeit

Aufgrund der deutlichen Trennung von (1) von den anderen Komponenten (> 7 min) bot sich der Wechsel zu einer kürzeren Säule als Option mit einer erheblichen Verkürzung der Retentionszeit an. Die ursprünglich 250 mm lange Säule wurde durch eine baugleiche Säule mit halber Länge (125 mm) ersetzt. Zusätzlich ist die Flussrate von 1,00 auf 1,25 mL/min erhöht worden (s. 6.4.1). Die Retentionszeit von (1) konnte so, wie in Abbildung 3.1-2 zu sehen, auf unter 10 min gesenkt werden.

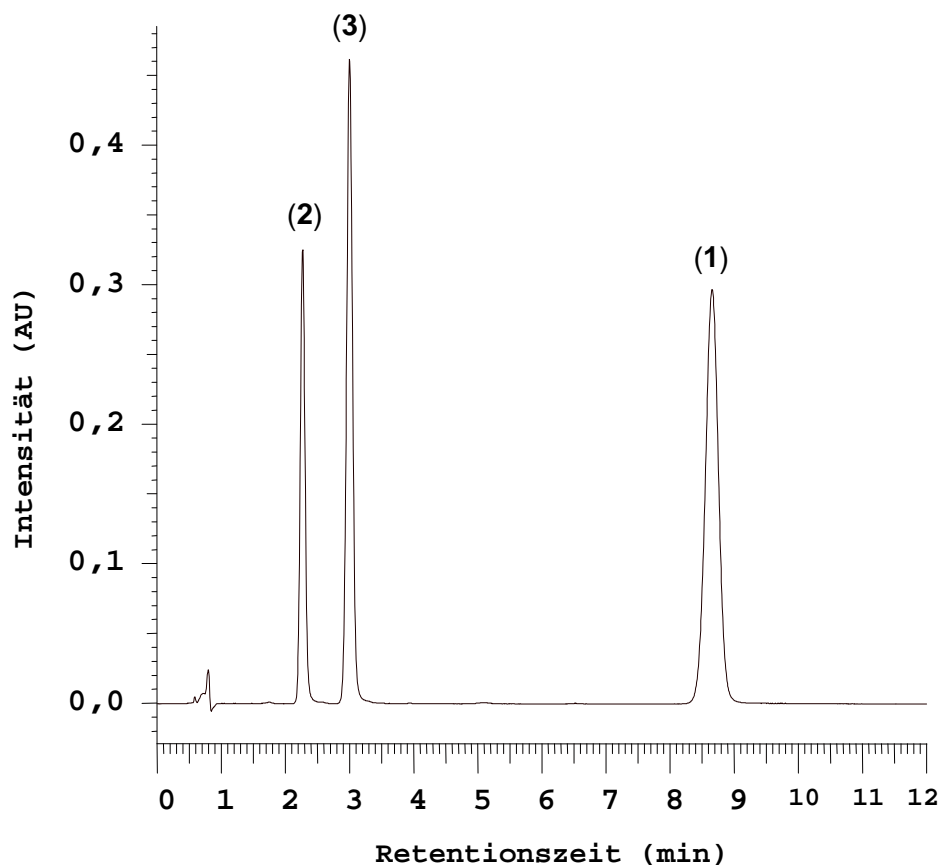


Abbildung 3.1-2: Die Retentionszeit von (1) sinkt unter Beibehaltung der Trennung auf unter 10 min.

3.1.2 Erhöhung der Stabilität der Probenlösung

Die Stabilität der ursprünglichen Probenlösung von (1) war bei 22 °C nur für ca. drei Stunden als stabil anzusehen, da sie sich hydrolytisch zu (2) zersetzt. Unter

der Annahme, dass sich die Stabilität durch den pH-Wert beeinflussen lässt, wurden drei verschiedene pH-Werte in einem Vorversuch daraufhin untersucht. Um einen eng umgrenzten pH-Bereich in der Probenlösung zu erhalten, erfolgte der Zusatz eines Phosphatpuffers zu den Lösungen. Für die pH-Werte 3, 5 und 7 wurden entsprechende Puffer hergestellt und getestet, indem in ca. viertelstündigen Abständen über mehrere Stunden hinweg die Vermessung verschiedener Proben erfolgte. In Abbildung 3.1-3 sind die Einzelwerte als prozentualer Anteil der Fläche an der Gesamtfläche der drei verschiedenen pH-Werte als Funktion der Zeit graphisch dargestellt.

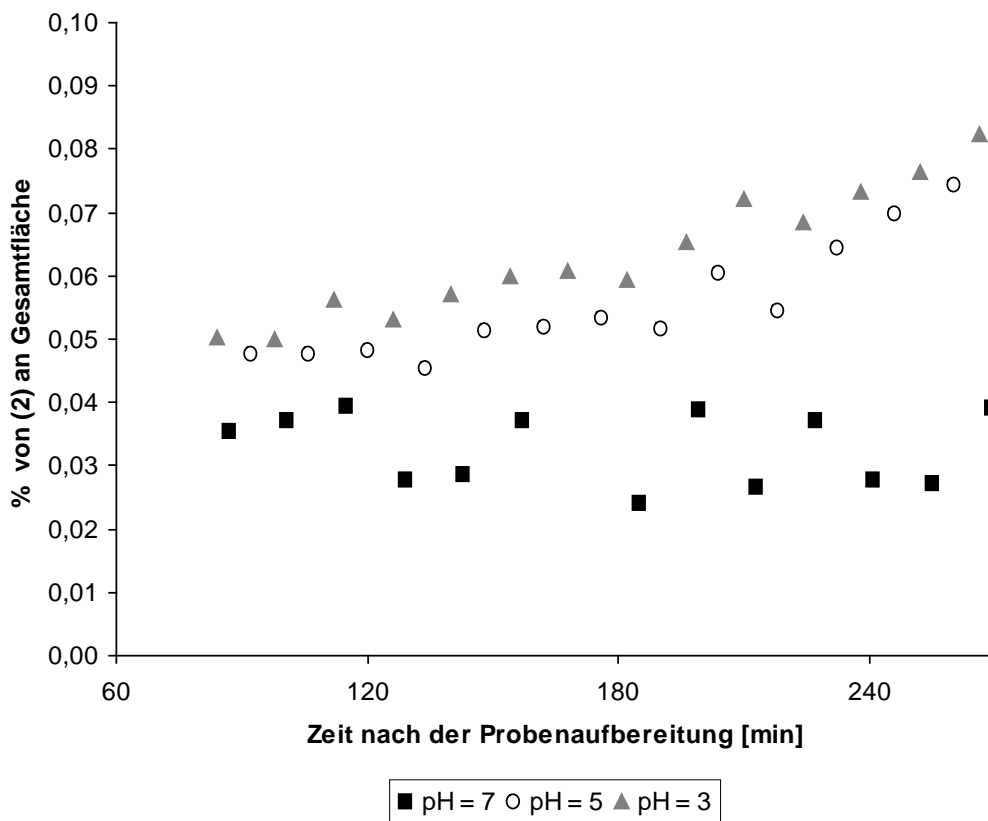


Abbildung 3.1-3: Darstellung der Einzelwerte der Peakflächen von (2) bezogen auf die Gesamtfläche (mit dem Hauptpeak (1)) in den ersten vier Stunden nach der Probenaufbereitung.

Der Einfachheit halber wurde aus den Messergebnissen jeweils anhand linearer Regression eine mittlere Bildungsgeschwindigkeit von (2) aus Glibenclamid ermittelt. Da hier von einem nichtlinearen Prozess ausgegangen werden kann, lassen sich die so erhaltenen Werte nur als eine Näherung auffassen.

Hierbei zeigte der pH-Wert 7 mit einem Wert von weniger als 0,01 %/Stunde die langsamste Bildungsgeschwindigkeit der drei Pufferlösungen, was als ausreichend für die Messserien angesehen wurde. Die Extrapolation der Regressionsgraden auf den Zeitpunkt 0 Minuten zeigt zudem, dass bereits in der Ausgangssubstanz von (1) ein kleiner Teil der Komponente (2) enthalten ist.

3.1.3 Säulenofen

Während der Vorserien wurde ein Säulenofen in das HPLC-System integriert. Dadurch sollte der Einfluss von Temperaturschwankungen zwischen den verschiedenen Analysetagen und innerhalb der jeweiligen Serien auf die Retentionszeit und damit auf die Peakflächen minimiert werden. Um nicht das ganze System temperieren zu müssen, wurde eine Ofentemperatur von 35 °C gewählt. Mit einer gegenüber der Raumtemperatur nur mäßig erhöhten Säulentemperatur kann eine Peakverbreiterung aufgrund eines Temperaturgradienten entlang der Säule umgangen werden [48].

Die Variation der Retentionszeiten innerhalb einer Serie wurde durch den Säulenofen von ca. 0,30 % auf 0,15 % halbiert. Die Systempräzision von anfänglich ca. 0,49 % verbesserte sich auf 0,31 % (s. 3.2.3).

3.1.4 Verringerung der Peakbreite bei Injektion von 100 µL (Reihe B)

Die Erhöhung der Empfindlichkeit der Methode in der zweiten von insgesamt drei Messreihen (Reihe B) wurde anhand eines 10-fach vergrößerten Injektionsvolumens erzielt, indem 100 µL injiziert wurden (vorher 10 µL). Bei Anwendung der Methode aus Reihe A trat anfangs eine starke Peakverbreiterung mit einer Peakbreite von mehr als 1,5 Minuten auf, die es zu verbessern galt (s. Abb. 3.1-4).

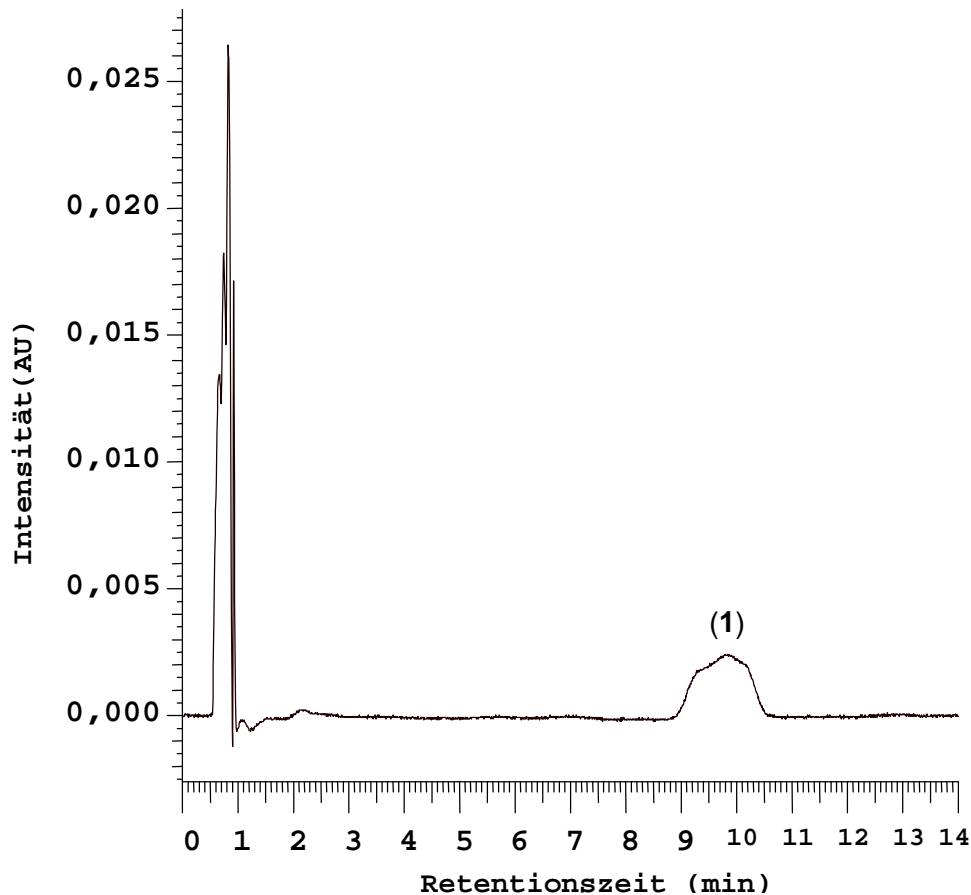


Abbildung 3.1-4: Peakverbreiterung von (1) nach Injektion von 100 μL . Die Breite und die ungünstige Peakform sollten optimiert werden.

Mehrere Versuche, mit Gradientenelution die Peakbreite zu minimieren, zeigten nicht den gewünschten Erfolg. Außerdem hätte die so geänderte Methode in den Chromatogrammen einen Versatz in der Basislinie nach sich gezogen.

Die effektivste und einfachste Problembehebung bestand letztlich in einer kleinen Abwandlung in der Probenaufbereitung: Die Änderung der Zusammensetzung des Lösungsmittels von Acetonitril-Puffer (80:20, v/v) zu Acetonitril-Puffer (40:60, v/v) zeigte den optimalen Effekt. Durch Injektion eines Mediums, in dem sich der Analyt weniger gut löst, ließen sich die Peakbreite (auf weniger als eine Minute) und die Peakform auf ein ausreichendes Maß verbessern (s. Abb. 3.1-5).

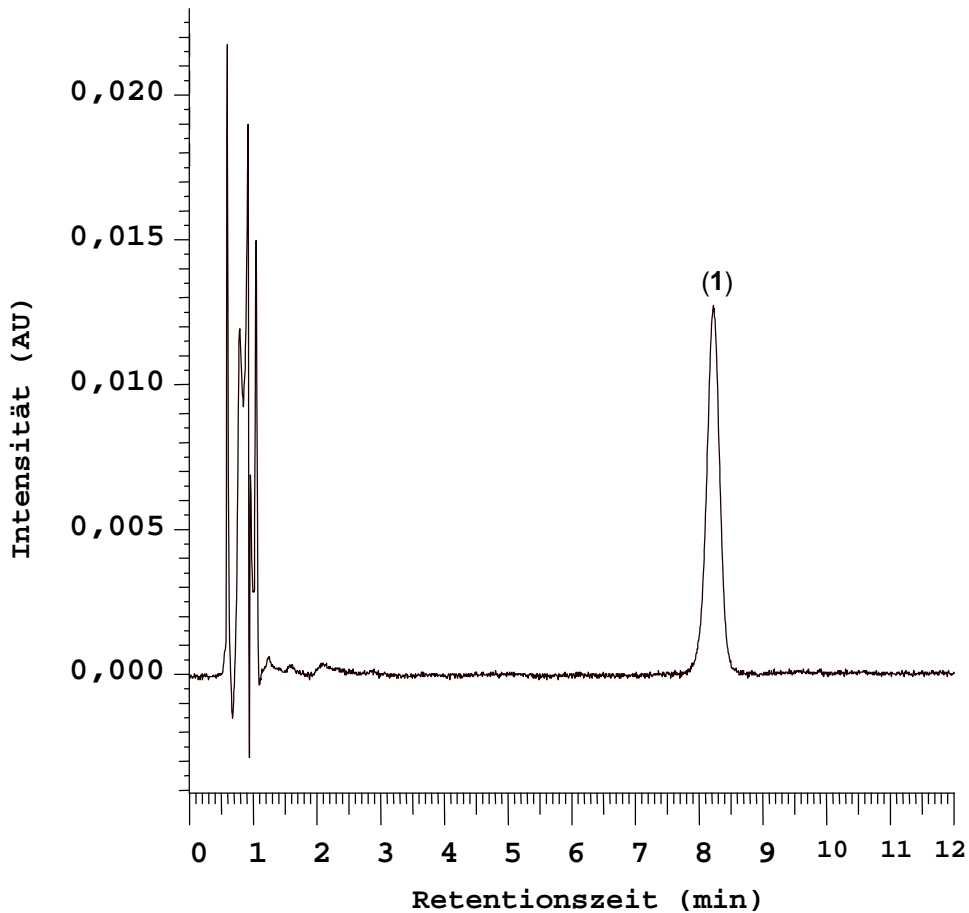


Abbildung 3.1-5: Die Injektion von 100 μL aus einer Probenlösung von (1) in Acetonitril-Puffer (40:60, v/v) zeigt eine deutliche Verringerung der Peakbreite und eine übliche Peakform.

3.1.5 Systemeignungstest

Nach Abschluss der Maßnahmen zur Methodenoptimierung (kürzere Säule, Probenstabilität, Säulenofen) wurde ein Systemeignungstest durchgeführt [49], indem eine Probe von (1) hergestellt und $n = 18$ mal anhand der Methode (s. 6.4.1) vermessen wurde. Die Bestimmung der Systempräzision ergab dabei eine RSD% von 0,5 %.

3.2 Ergebnisse zur Präzision

Die Untersuchungsziele dieser Arbeit hinsichtlich der Präzisionsmaße Wiederhol-Standardabweichung ($\hat{\sigma}_r$) und Systempräzision ($\hat{\sigma}_{sys}$) in der HPLC sind anhand von drei Messreihen bearbeitet worden. Diese Messreihen werden im weiteren Verlauf als A, B und C bezeichnet und sind in Schema 3.2-1 abgebildet. Sie bestanden jeweils aus mehreren Serien, die an verschiedenen Tagen durchgeführt wurden. Alle Standards und alle Proben sind voneinander unabhängig hergestellt worden.

Reihe A	Reihe B	Reihe C
Gehaltsbestimmung Arzneiform	Gehaltsbestimmung Arzneiform Erhöhte Empfindlichkeit	Gehaltsbestimmung Arzneistoff
Glibenclamid Matrix	Glibenclamid Matrix	Glibenclamid (3 - 6)

Schema 3.2-1: Pensum der HPLC Untersuchungen.

In Reihe A ist Glibenclamid in mehreren Serien über einen weiten Konzentrationsbereich von 120 bis 0,025 % anhand ein und derselben Methode bestimmt worden, wobei das Konzentrationsniveau 100 % einer tatsächlichen Konzentration von 0,2 mg/mL entsprach. Um die Situation einer routinemäßigen Gehaltsbestimmung einer Arzneiform nachzustellen, wurde (1) neben einer konstanten Menge einer üblichen Tablettenmatrix quantifiziert (s. 6.7.3).

In Reihe B sollte der Einfluss einer im Vergleich zu Reihe A empfindlicheren Methode auf die $\hat{\sigma}_r$ untersucht werden. Dazu wurde ein Konzentrationsbereich von 0,1 bis 0,01 % gewählt, so dass sich die Konzentrationsniveaus 0,1 und 0,05 % in beiden Reihen wiederfanden.

Um eine Reinheitskontrolle eines Arzneistoffes nachzustellen, wurden in Reihe C neben einer festen Konzentration des Glibenclamids von 100 % die Substanzen (3-6) als Modellverunreinigungen zugesetzt (s. 6.7.8).

In den verschiedenen Serien ist die Zusammensetzung der Konzentrationen der einzelnen Verunreinigungen zufällig angeordnet worden (s. Tab. 6.7-5). Jede Verunreinigung wurde letztlich über einen Konzentrationsbereich von 1,0 bis 0,025 % vermessen. Da sich die Substanzen in ihren chromatographischen Eigenschaften voneinander unterscheiden, sollte so versucht werden, den Einfluss der Eigenschaften wie Retentionszeit oder spezifische Absorption auf den Verlauf der Präzision zu untersuchen.

Zusätzlich existieren in dieser Reihe verschiedene Trennsituationen zwischen den einzelnen Verbindungen wie vollständig basisliniengetrennte, nicht mehr ganz basisliniengetrennte und angetrennte Peaks, so dass die Effekte unterschiedlicher Trennsituationen auf den Verlauf der Präzision beobachtet werden können.

3.2.1 Beziehung zwischen der Präzision und der Konzentration

3.2.1.1 Ergebnisse der Reihe A

Im Rahmen der Reihe A wurden die Konzentrationsniveaus der einzelnen Serien anhand des Probenaufbereitungsschemas aus Abschnitt 6.7.8 ausgehend vom Konzentrationsniveau 120 % schrittweise verringert, bis das Signal-Rausch-Verhältnis S/N den Wert 2 erreicht hatte (S/N bestimmt nach [23]). Das tiefste Konzentrationsniveau lag somit unter der Nachweis- und Bestimmungsgrenze ($S/N > 3$, bzw. $S/N > 10$ oder Präzision der Bestimmung unter 10 %). Dies war bei einer nominalen Konzentration von 0,025 % der Fall. Abbildung 3.2-1 zeigt ein Chromatogramm der Konzentration 100 % aus der Reihe A.

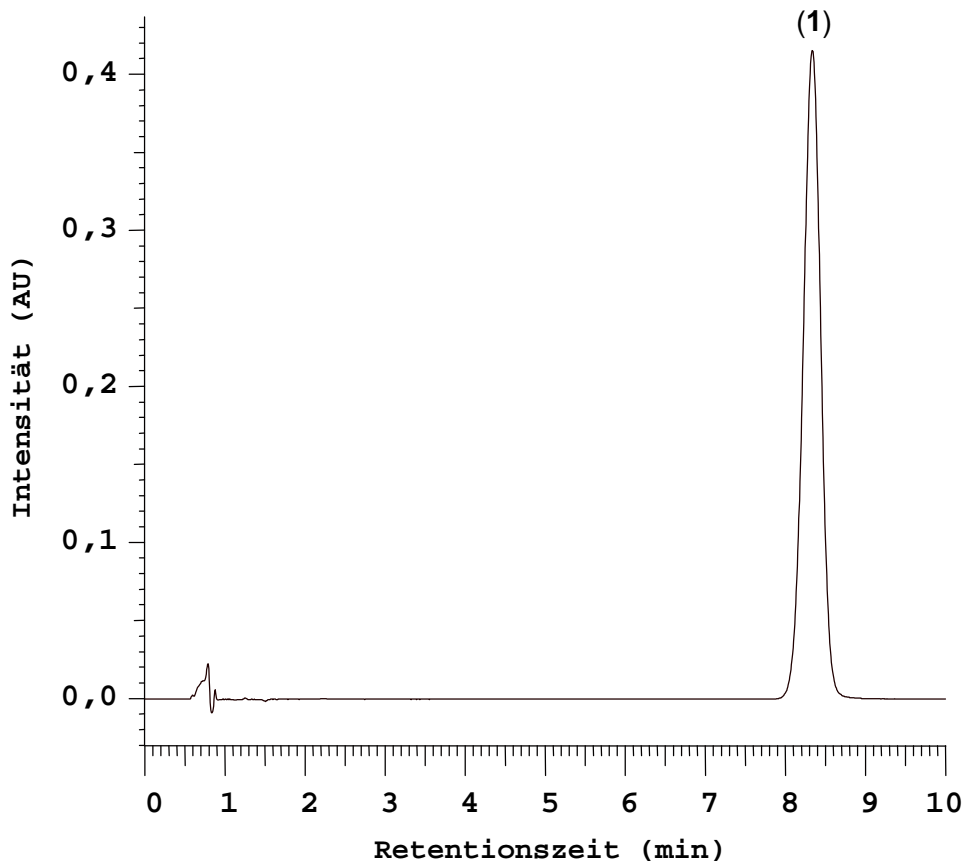


Abbildung 3.2-1: Beispielchromatogramm aus Reihe A. Die Probenkonzentration betrug 0,2 mg/mL bzw. 100 %.

Die aus den Serien berechneten Werte für die Präzisionsmaße $\hat{\sigma}_r$ und $\hat{\sigma}_{sys}$ sind in der nachfolgenden Tabelle 3.2-1 aufgeführt.

Konzentrations-niveau [%]	$\hat{\sigma}_r$ [%]	$\hat{\sigma}_{sys}$ [%]
120	0,60	-
100	0,57	0,35
80	0,77	0,38
60	0,55	0,32
40	0,46	0,26
20	1,09	0,31
10	0,45	0,26
7,5	1,93	0,28
7,5	0,63	0,29
5,0	1,25	0,42
2,5	0,62	0,75
1,0	1,32	1,24
0,75	2,27	1,30
0,50	1,89	1,46
0,25	4,89	1,70
0,10	3,71	3,42
0,075	4,35	9,47
0,050	13,69	8,13
0,025	15,14	14,64

Tabelle 3.2-1: RSD%-Werte für die $\hat{\sigma}_r$ und die $\hat{\sigma}_{sys}$ in Reihe A.

Abbildung 3.2-2 zeigt die RSD%-Werte des ganzen Bereichs als Überblick in logarithmischer Skalierung der x-Achse. Zu beachten ist, dass das Konzentrationsniveau 7,5 % zweimal vermessen wurde, da sich an dieser Stelle die Probenaufbereitung leicht geändert hatte. In den darauffolgenden Konzentrationen wurden andere Messkolbenvolumina und ein weiterer Verdünnungsschritt miteingefügt (s. Tab. 6.7-2).

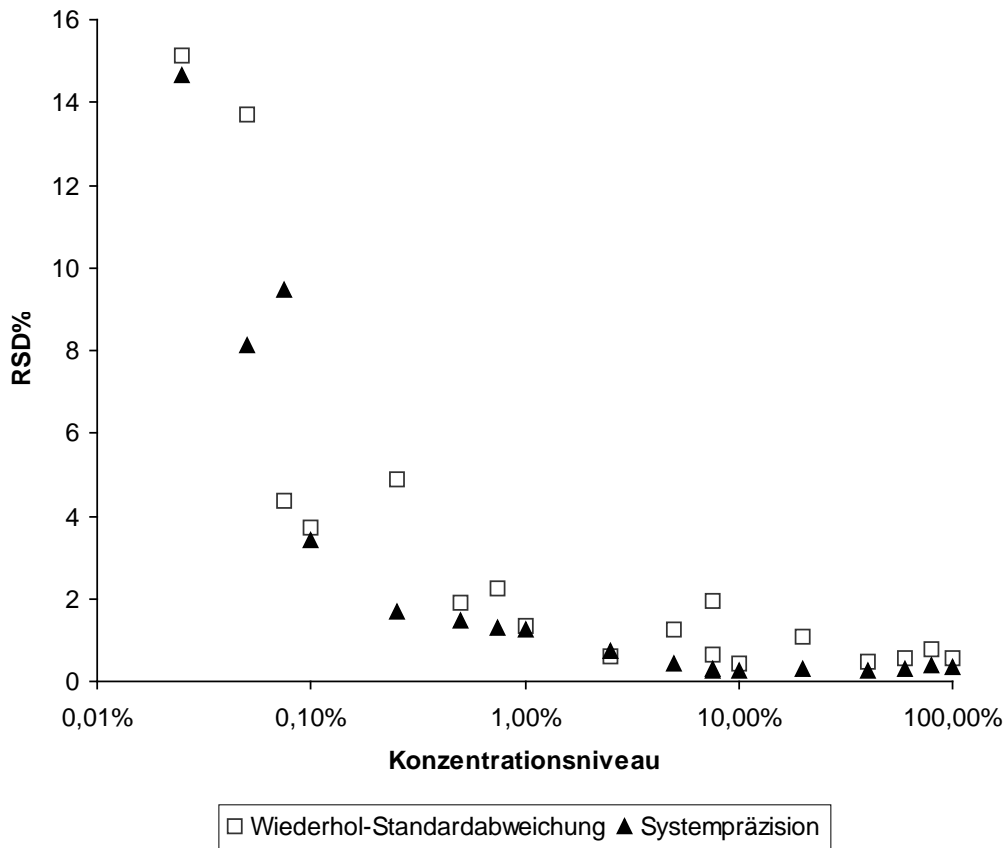


Abbildung 3.2-2: RSD%-Werte für $\hat{\sigma}_r$ und $\hat{\sigma}_{sys}$ über den ganzen Konzentrationsbereich in logarithmischer Darstellung der x-Achse.

Bezüglich des Verlaufs der $\hat{\sigma}_r$ und der $\hat{\sigma}_{sys}$ gegen kleine Konzentrationen lässt sich der ganze Konzentrationsbereich in drei Abschnitte gliedern, in denen bestimmte, verschiedene Zusammenhänge vorliegen, die eine Reihe von Schlussfolgerungen erlauben. Der erste Abschnitt deckt die Konzentrationen des Bereichs von 120 bis 10 %, der zweite den Bereich von 7,5 bis 0,25 % und der dritte den Bereich von 0,1 bis 0,025 % ab.

Im Bereich höherer Konzentrationen von 120 bis 10 % ist zu erkennen, dass sich die $\hat{\sigma}_r$ und die $\hat{\sigma}_{sys}$ (von kleinen Schwankungen abgesehen) konstant verhalten, während die Konzentration in diesem Bereich auf unter 1/10 reduziert worden ist (s. Abb. 3.2-3).

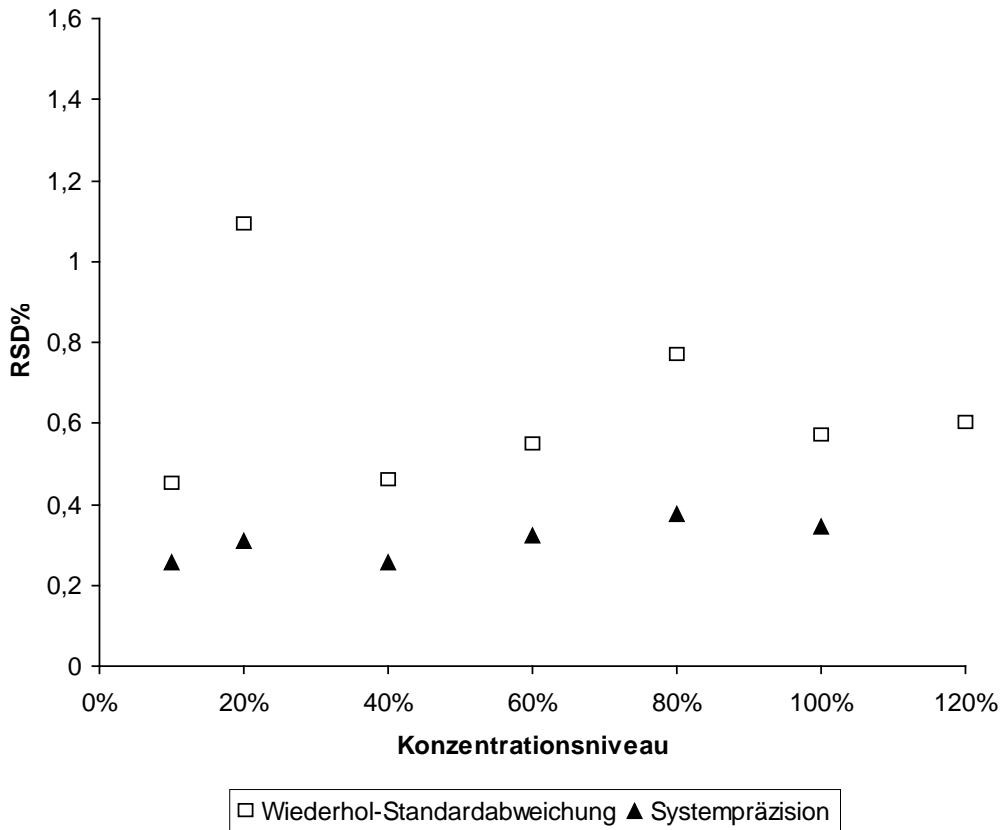


Abbildung 3.2-3: RSD%-Werte für $\hat{\sigma}_r$ und $\hat{\sigma}_{sys}$ für die höheren Konzentrationen von 120 bis 10 %. Sowohl $\hat{\sigma}_r$ als auch $\hat{\sigma}_{sys}$ verlaufen nahezu konstant.

Der paarweise Vergleich von $\hat{\sigma}_r$ und $\hat{\sigma}_{sys}$ innerhalb eines Konzentrationsniveaus anhand des F-Testes im Bereich zwischen 10 und 100 % ergibt nicht immer einen signifikanten Unterschied. Dies ist auf die geringe Anzahl an Freiheitsgraden ($n_{1-6} = 6$) zurückzuführen. Die Wahrscheinlichkeit des F-Testes im Erkennen einer statistischen Signifikanz steigt mit zunehmender Stichprobengröße an (s. 2.6.7).

Unter der Annahme, dass die einzelnen Bestimmungen der $\hat{\sigma}_r$ und der $\hat{\sigma}_{sys}$ bei den verschiedenen Konzentrationen in diesem Bereich unabhängige Stichproben aus jeweils derselben Grundgesamtheit $N(\mu, \sigma_r^2)$ bzw. $N(\mu, \sigma_{sys}^2)$ sind, können die Stichproben zu einer gemeinsamen Messreihe nach Gleichung 3.2-1 vereinigt werden [32]:

$$\hat{\sigma}_p^2 = \frac{(n_1 - 1)\hat{\sigma}_1^2 + \dots + (n_6 - 1)\hat{\sigma}_6^2}{(n_1 - 1) + \dots + (n_6 - 1)} \quad (3.2-1)$$

Die so gepoolten Werte für die $\hat{\sigma}_r$ und die $\hat{\sigma}_{sys}$ ergeben RSD%-Werte von 0,70 % bzw. 0,31 %. Berechnet man den F-Test mit den vereinigten Werten und je 30 Freiheitsgraden und einem $\alpha = 0,05$, erhält man einen Prüfwert T_F von 2,26, der größer ist als der kritische Wert 1,36. Daher lässt sich die Probenaufbereitung als signifikante Fehlerquelle in diesem Bereich auffassen, und zwar als die größte (s. 3.2.2).

Im darauf folgenden Bereich von 7,5 bis 0,25 % (s. Abb. 3.2-4) zeigt die $\hat{\sigma}_{\text{sys}}$ gegenüber der logarithmierten Konzentration einen linear ansteigenden Verlauf, ähnlich der Horwitz-Beziehung. Horwitz hat aus der Betrachtung einer großen Anzahl unterschiedlicher Methoden und Konzentrationen einen allgemeinen Zusammenhang zwischen Präzision und der Konzentration des Analyten in der Ausgangssubstanz abgeleitet [50, 51, 52]. Die Beziehung besagt, dass sich unabhängig von der Art der Methode der RSD%-Wert der Präzision (als Ringversuchs-Standardabweichung in RSD%) verdoppelt, wenn die Konzentration auf 1/100 sinkt. Diese Beziehung ergibt bei logarithmischer Betrachtung einen linearen Verlauf.

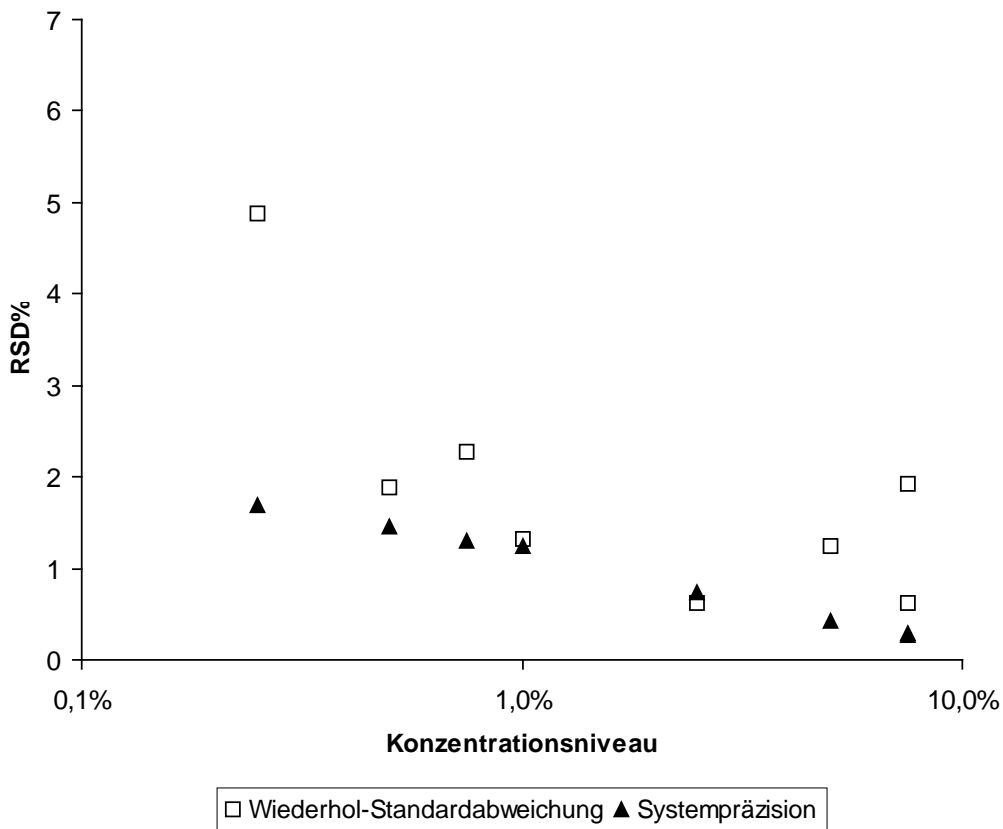


Abbildung 3.2-4: Darstellung der $\hat{\sigma}_r$ und der $\hat{\sigma}_{\text{sys}}$ zwischen 7,5 und 0,25 %. Die $\hat{\sigma}_{\text{sys}}$ zeigt einen lineare Anstieg gegenüber dem Logarithmus der Konzentration.

Die $\hat{\sigma}_r$ ist im Unterschied zur $\hat{\sigma}_{\text{sys}}$ in diesem Bereich durch die zusätzliche Streuung der Probenaufbereitung stärkeren Schwankungen unterworfen, wodurch das Vorliegen eines linearen Verhaltens nur schwer abzulesen ist.

Unterhalb von 0,25 % geht der Verlauf in einen nichtlinearen Zusammenhang über. Im Bereich unterhalb von 0,25 % erfolgt in Abbildung 3.2-5 ein äußerst starker Anstieg der RSD%-Werte für die $\hat{\sigma}_r$ und die $\hat{\sigma}_{sys}$.

Die Werte für die $\hat{\sigma}_r$ und die $\hat{\sigma}_{sys}$ stehen in den letzten drei Konzentrationsstufen in keiner klaren Beziehung mehr zueinander. Während bei 0,025 % beide Präzisionsmaße eng zusammen liegen, unterscheiden sich die $\hat{\sigma}_r$ und die $\hat{\sigma}_{sys}$ bei 0,050 % stark. Bei 0,075 % übersteigt aufgrund der hohen Streuung die $\hat{\sigma}_{sys}$ betragsmäßig sogar die $\hat{\sigma}_r$.

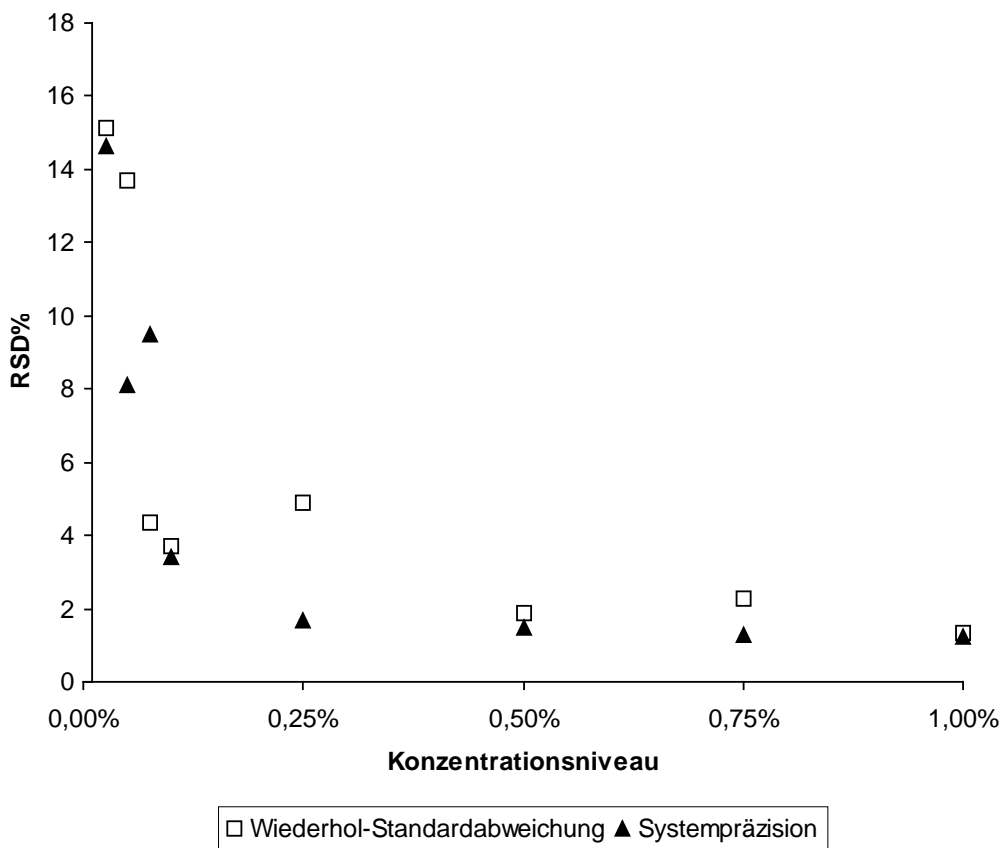


Abbildung 3.2-5: Der unterste Konzentrationsbereich. Die RSD%-Werte beider Präzisionsmaße steigen gleichermaßen stark an.

Für den Konzentrationsbereich von 1,0 bis 0,025 % ist die Systempräzision $\hat{\sigma}_{sys}$ zusätzlich mit der empfindlicheren Detektoreinstellung „low mode“ des L 7450 Diodenarraydetektors vermessen worden, um herauszufinden, ob die beiden Modi zu unterschiedlichen Präzisionen führen. Beim „low mode“ lässt sich im Vergleich

zum sogenannten „normal mode“ eine 10-fach erhöhte Signalauflösung erhalten, indem als kleinste Absorptionseinheit 0,005 mAU statt 0,05 mAU ausgegeben wird.

Die so mit dem „low mode“ bestimmten RSD%-Werte für die $\hat{\sigma}_{\text{sys}}$ dieses Konzentrationsbereiches sind in Abbildung 3.2-6 gemeinsam mit den RSD%-Werten des „normal mode“ dargestellt. Daraus wird ersichtlich, dass die Umschaltung auf den „low mode“ keinen nennenswerten Einfluss auf die Systempräzision besitzt.

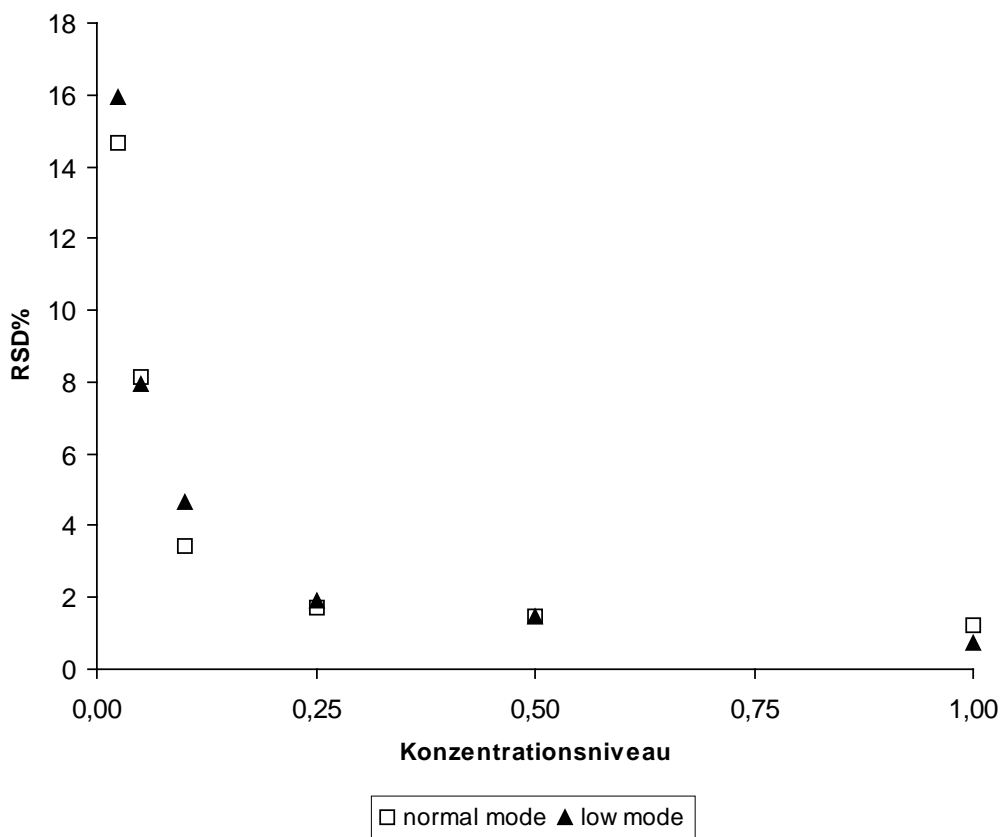


Abbildung 3.2-6: Ein Unterschied in den Systempräzisionen zwischen den Detektoreinstellungen „normal mode“ und „low mode“ ist nicht zu erkennen.

Im unteren Konzentrationsbereich von 0,5 bis 0,025 % unterscheiden sich die Werte für $\hat{\sigma}_r$ und $\hat{\sigma}_{\text{sys}}$ aufgrund der aufgeweiteten Vertrauensbereiche nicht signifikant voneinander. Für jeweils $n-1 = 5$ Freiheitsgrade (bei 6 Messungen) und einer gewählten Irrtumswahrscheinlichkeit α von 0,05 müsste das Verhältnis der RSD%-Werte im Rahmen eines F-Testes den kritischen Wert 2,24 überschreiten,

um eine signifikante Differenz zu zeigen. Es ergibt sich bei einer Konzentration von z.B. 0,5 % mit den Werten 1,89 % für die $\hat{\sigma}_r$ und 1,46 % für die $\hat{\sigma}_{sys}$ jedoch nur der Wert 1,29.

Als Bestimmungsgrenze wurde diejenige Konzentration festgelegt, die sich mit einer RSD < 10 % noch bestimmen lässt [53]. Sie liegt bei 0,05 % (s. Tab. 3.2-1). Das Signal-Rausch-Verhältnis betrug an dieser Stelle 6 (bestimmt nach [23]).

3.2.1.2 Ergebnisse der Reihe B

Die Erhöhung der Empfindlichkeit um den Faktor 10 erfolgte durch eine Erhöhung des Injektionsvolumens von 10 auf 100 μ L. Eine dadurch anfänglich auftretende Peakverbreiterung konnte beseitigt werden, indem die Konzentration des Acetonitrils in der Probenlösung von 80 auf 40 % halbiert wurde. Die aus den Serien berechneten Werte für die $\hat{\sigma}_r$ und die $\hat{\sigma}_{sys}$ sind in der nachfolgenden Tabelle 3.2-2 aufgeführt.

	Reihe A	Reihe B	Reihe B
Konzentrations-niveau [%]	$\hat{\sigma}_r$ [%]	$\hat{\sigma}_r$ [%]	$\hat{\sigma}_{sys}$ [%]
0,50	1,89	-	-
0,10	3,71	1,33	1,39
0,050	13,69	2,02	1,51
0,010	-	5,06	3,86

Tabelle 3.2-2: Werte für die $\hat{\sigma}_r$ aus Reihe B und A (als Vergleich) und die entsprechenden Werte für die $\hat{\sigma}_{sys}$ der Reihe B.

Für die Konzentrationen der Reihe B sind Doppelinjektionen (s. 6.6) durchgeführt worden. Daher lässt sich die Systempräzision unter Verwendung nachfolgender Gleichung 3.2-2 berechnen (m steht für die Anzahl an Doppelinjektionen).

$$\hat{\sigma}_{sys} = \sqrt{\frac{\sum_{i=1}^m (x_{i,1} - x_{i,2})^2}{2m}} \quad (3.2-2)$$

Abbildung 3.2-7 zeigt die beiden Messreihen A und B in einem überlappenden Konzentrationsbereich.

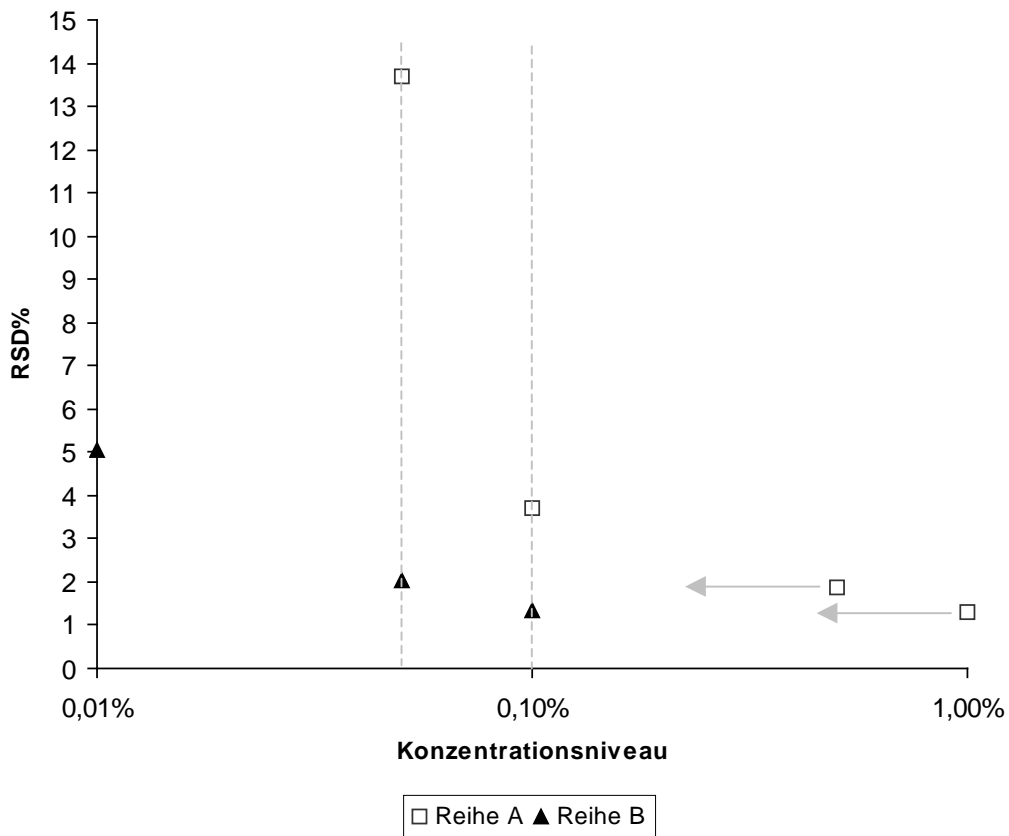


Abbildung 3.2-7: RSD%-Werte der $\hat{\sigma}_r$ von Reihe A und B des überlappenden Konzentrationsbereichs. Die $\hat{\sigma}_r$ bei 1,0 % in Reihe A entspricht betragsmäßig in etwa der $\hat{\sigma}_r$ an der Stelle 0,10 % in Reihe B. Die Konzentration von 0,01 % war in Reihe A nicht mehr bestimmbar.

Es zeigt sich eine deutliche Verbesserung der $\hat{\sigma}_r$ in Reihe B (s. gestrichelte Linien). Vergleicht man die Werte der beiden Reihen, ist zu sehen, dass die $\hat{\sigma}_r$ einer Konzentration der Reihe B der $\hat{\sigma}_r$ aus der Reihe A bei einer 10-fach höheren Konzentration gleicht (s. Pfeile in Abb. 3.2-7). Da die Empfindlichkeit auch um den Faktor 10 erhöht wurde, lässt sich hieraus schlussfolgern, dass in diesen Konzentrationsbereichen die Systempräzision, die durch die Erhöhung der Empfindlichkeit beeinflusst wurde, bereits die absolut dominierende Komponente an der Gesamtvariation ist.

3.2.1.3 Ergebnisse der Reihe C

Abbildung 3.2-8 zeigt ein typisches Chromatogramm dieser Reihe mit (1) als Hauptkomponente neben (3-6) als Verunreinigungen (s. 6.2). Die einzelnen Substanzen liegen in verschiedenen Trennsituationen vor: So ist (3) von allen anderen Substanzen deutlich getrennt. Zwischen (4) und (6) kann mit einer Auflösung von 1,8 noch von einer Basislinientrennung gesprochen werden. Nebenprodukt (5) hingegen ist von dem sich stets bildenden Abbauprodukt (2) nicht mehr basisliniengetreunt.

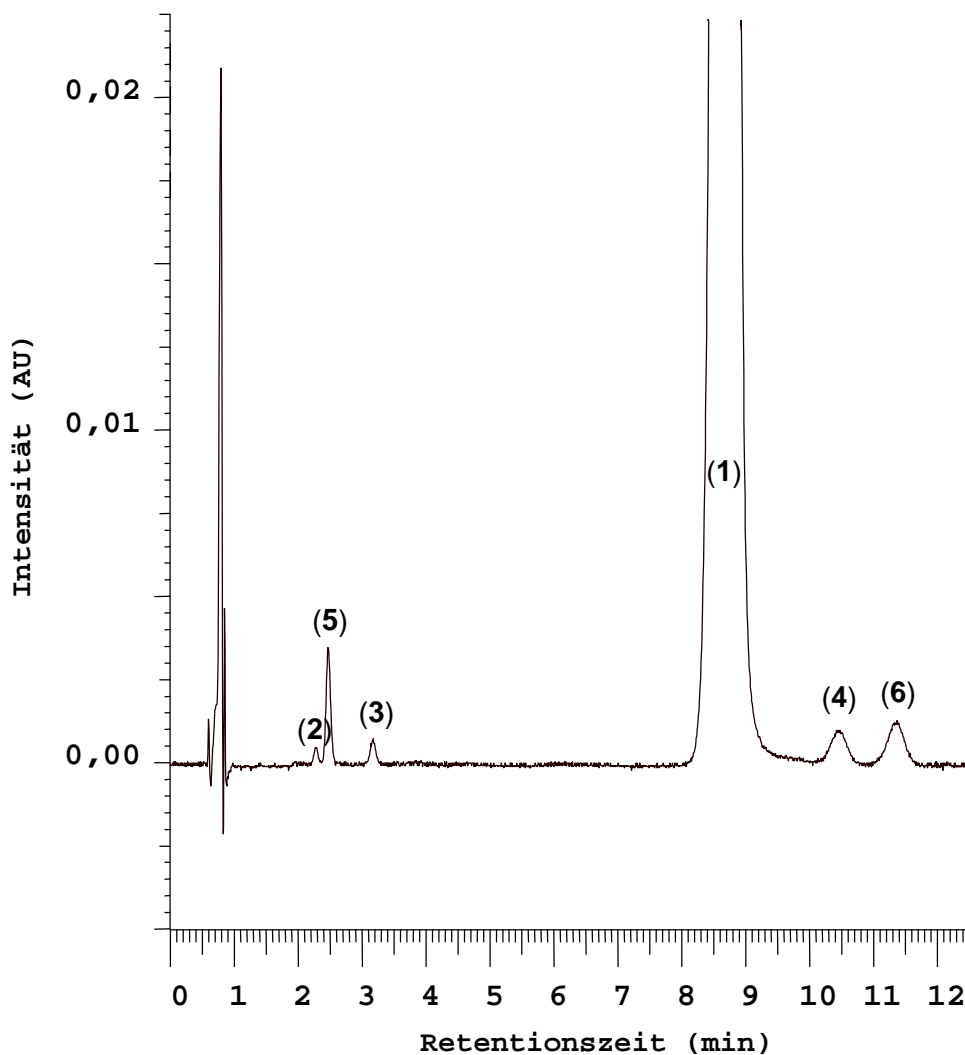


Abbildung 3.2-8: Zu sehen sind die Modellverunreinigungen (3-6) neben dem Hauptpeak des Glibenclamids (1).

Grundsätzlich zeigen die NebenkompONENTEN das gleiche Verhalten bezüglich $\hat{\sigma}_r$ und $\hat{\sigma}_{sys}$ wie Glibenclamid. Hinsichtlich des Niveaus, von dem aus die Variation sehr stark zunimmt, bestehen Unterschiede zwischen den Substanzen. In Abbildung 3.2-9 sind die Wiederhol-Standardabweichungen aller Substanzen in Abhängigkeit der Konzentration aufgetragen. Zum besseren Vergleich sind auch die Werte von Glibenclamid aus Reihe A abgebildet. Die Unterschiede in der $\hat{\sigma}_r$ lassen sich anhand der UV-Absorption, der Detektionswellenlänge und des zugehörigen Rauschens sowie mit der Retentionszeit erklären.

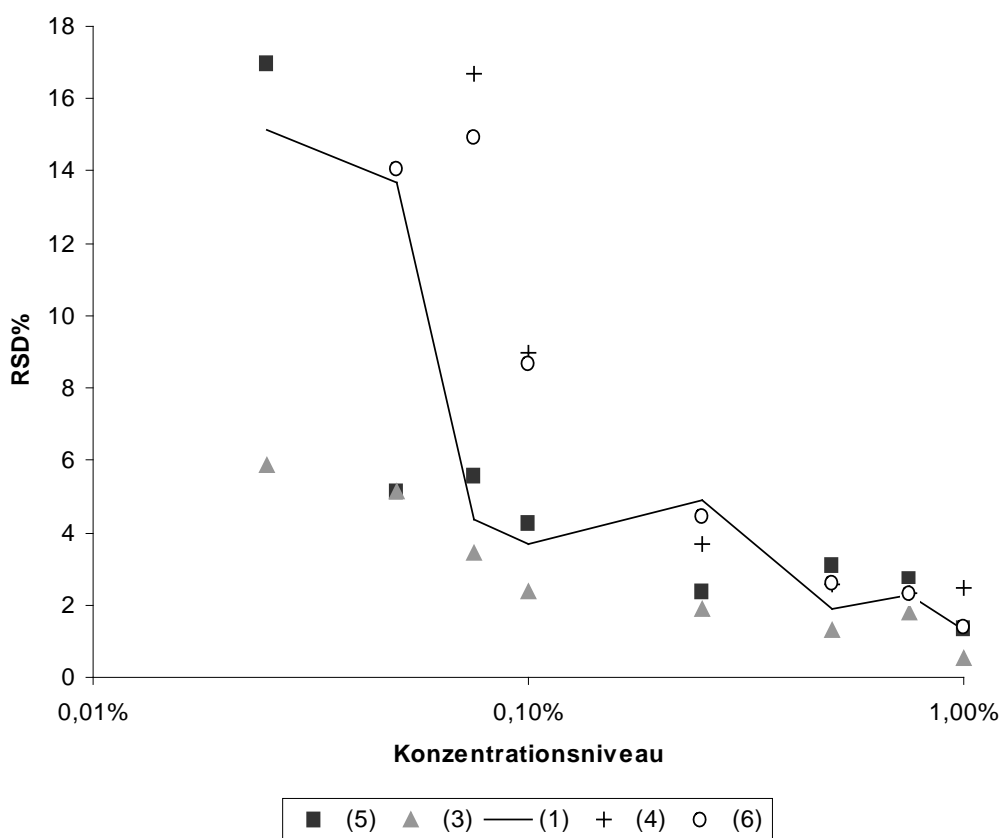


Abbildung 3.2-9: RSD%-Werte für die Wiederhol-Standardabweichungen der Modellverunreinigungen (3-6) aus Reihe C und für Glibenclamid aus Reihe A.

Der hohe Einfluss der UV-Absorption auf den Integrationsfehler lässt sich durch einen Vergleich der RSD%-Werte von (1) und (6) aus der Tabelle 3.2-3 aufzeigen. Die beiden Substanzen weisen eine ähnliche Retentionszeit auf und werden bei gleicher Wellenlänge detektiert; (1) besitzt allerdings nur die halbe UV-Absorption im Vergleich zu (6) (s. Abb. 3.2-10).

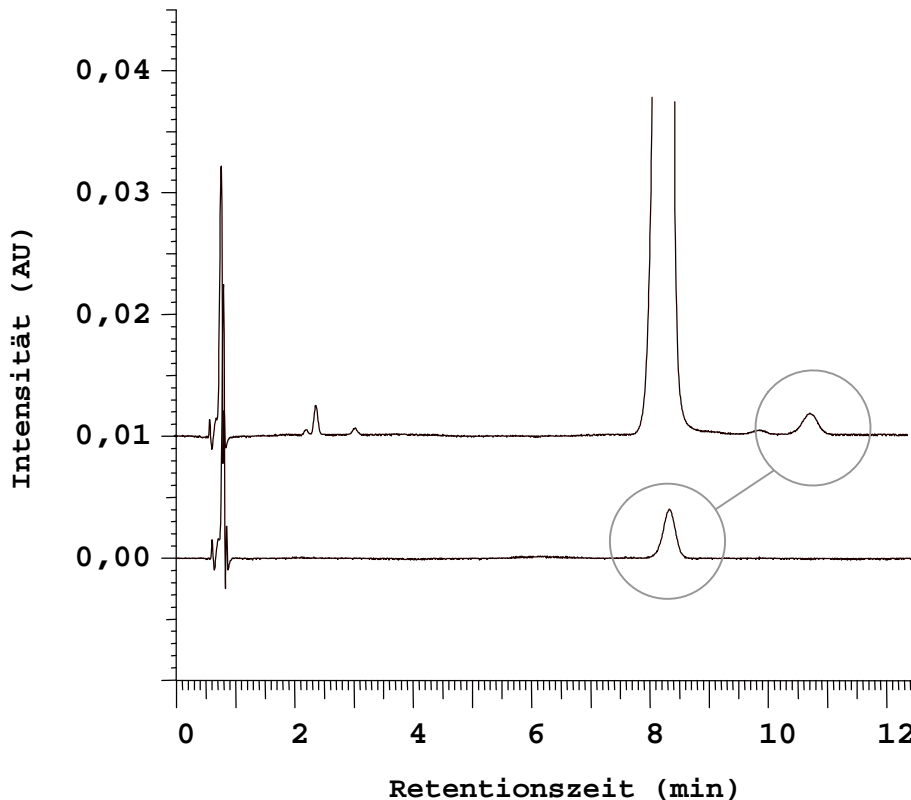


Abbildung 3.2-10: Die Abbildung zeigt (1) bei 8,25 Minuten und (6) bei 10,80 Minuten als markierte Peaks jeweils in einer Konzentration von 1,0 % aus zwei unterschiedlichen Läufen.

Durch den Vergleich der RSD%-Werte von (4) und (6) - im gleichen Chromatogramm detektiert bei 228 bzw. 210 nm - lässt sich der Effekt der Detektionswellenlänge auf die Systempräzision zeigen. Die Retentionszeit sowie die UV-Absorption sind ähnlich. Dieser beobachtete Unterschied in der Systempräzision (s. Tab. 3.2-3, Zeilen 8 und 10) ist wahrscheinlich durch das fast doppelte Rauschen bei 228 nm (verglichen mit 210 nm) entstanden.

		Konzentrationsniveau [%]							
		1,0	0,75	0,50	0,25	0,10	0,075	0,050	0,025
$\hat{\sigma}_r$	(1)	1,32	2,27	1,89	4,89	3,71	4,35	13,69	15,14
	(3)	0,52	1,79	1,32	1,87	2,37	3,44	5,12	5,89
	(4)	2,48	2,34	2,58	3,67	8,97	16,67	n.b.	n.b.
	(5)	1,34	2,73	3,10	2,39	4,29	5,60	5,16	16,98
	(6)	1,42	2,31	2,61	4,45	8,67	14,93	14,09	n.b.
$\hat{\sigma}_{sys}$	(1)	1,24	1,30	1,46	1,70	3,42	9,47	8,13	14,64
	(3)	0,45	0,80	0,79	1,33	1,85	3,46	4,57	4,71
	(4)	1,18	2,47	2,46	3,31	7,66	12,86	n.b.	n.b.
	(5)	0,74	2,07	2,68	1,84	4,36	3,48	3,73	15,94
	(6)	0,60	1,66	2,12	4,37	6,15	9,98	16,44	n.b.

Tabelle 3.2-3: $\hat{\sigma}_r$ und $\hat{\sigma}_{sys}$ der NebenkompONENTEN (3-6) aus Reihe C und von (1) aus Reihe A. Von (4) und (6) konnten einige Konzentrationen nicht mehr bestimmt werden (n.b.).

Ein Vergleich der RSD%-Werte von (1) und (3) zeigt den ausgeprägten Einfluss der Retentionszeit auf die $\hat{\sigma}_{sys}$. Die mit 3,1 Minuten klar geringere Retentionszeit von (3) verglichen mit den 8,5 Minuten von (1) führt durch die günstigere Peakhöhe (s. Abb. 3.2-8) zu einer deutlichen Verringerung der RSD%-Werte für die $\hat{\sigma}_{sys}$ (z.B. 4,57 % an der Stelle 0,050 % für (3) gegenüber 8,13 % für (1)). Die Situation, dass (5) von (2) nicht mehr basisliniengetrennt ist, wurde aus dem Grund gewählt, um auch Effekte solcher Trennproblematiken auf die Präzision mitzuerfassen. Dies ist möglicherweise der Grund für höhere RSD-Werte bei (5) verglichen mit (3). Allerdings sind auch die anderen Parameter leicht unterschiedlich, so dass der zusätzliche Effekt auf die $\hat{\sigma}_{sys}$ aufgrund der Peakbenachbarung nicht gesichert ist.

3.2.2 Beziehung zwischen der Systempräzision und der Detektionswellenlänge

Des Weiteren wurde die $\hat{\sigma}_{\text{sys}}$ von (1) anhand der Wellenlängen 210, 228 und 240 nm bestimmt. Die geringere UV-Absorption von 228 nm verglichen zu 240 nm ist in Abbildung 3.2-11 zu sehen.

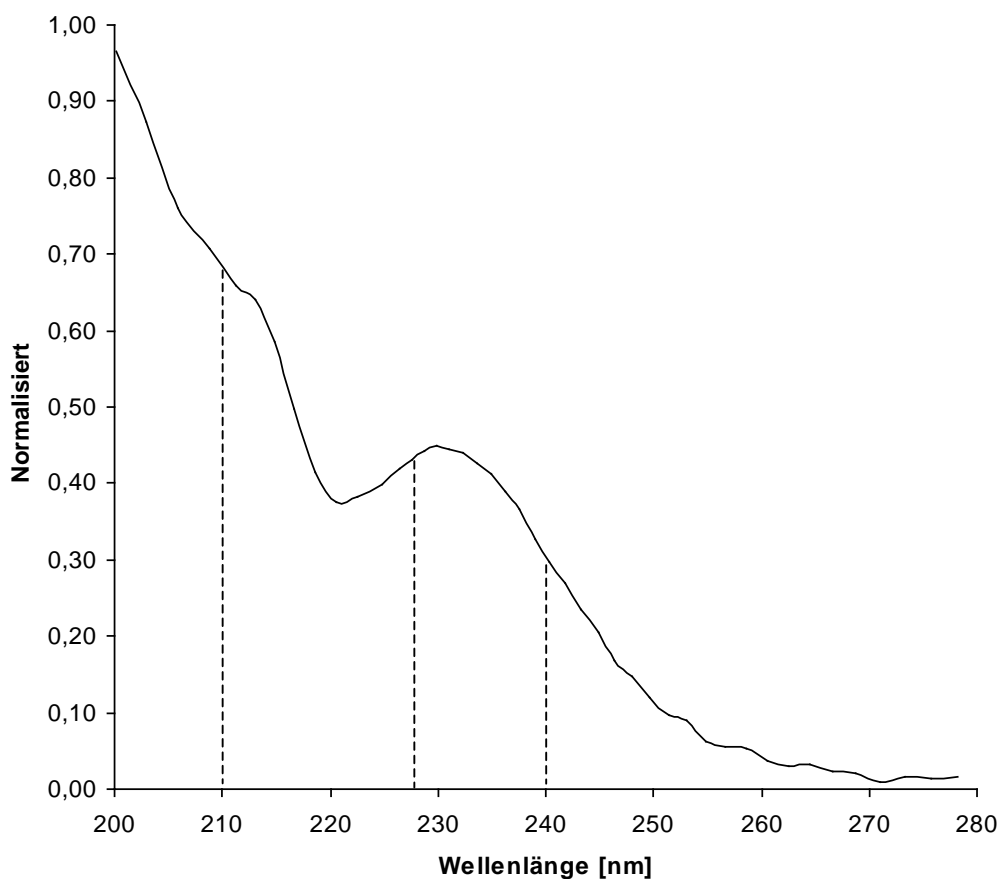


Abbildung 3.2-11: Das UV-Spektrum von (1), anhand der DAD Daten während der HPLC Experimente gewonnen. Gestrichelte Linien: Die beobachteten Wellenlängen 210, 228 and 240 nm.

Das Rauschen war in diesem Bereich unabhängig von der Wellenlänge. Der Einfluss auf die $\hat{\sigma}_{\text{sys}}$ ist in Abbildung 3.2-12 dargestellt.

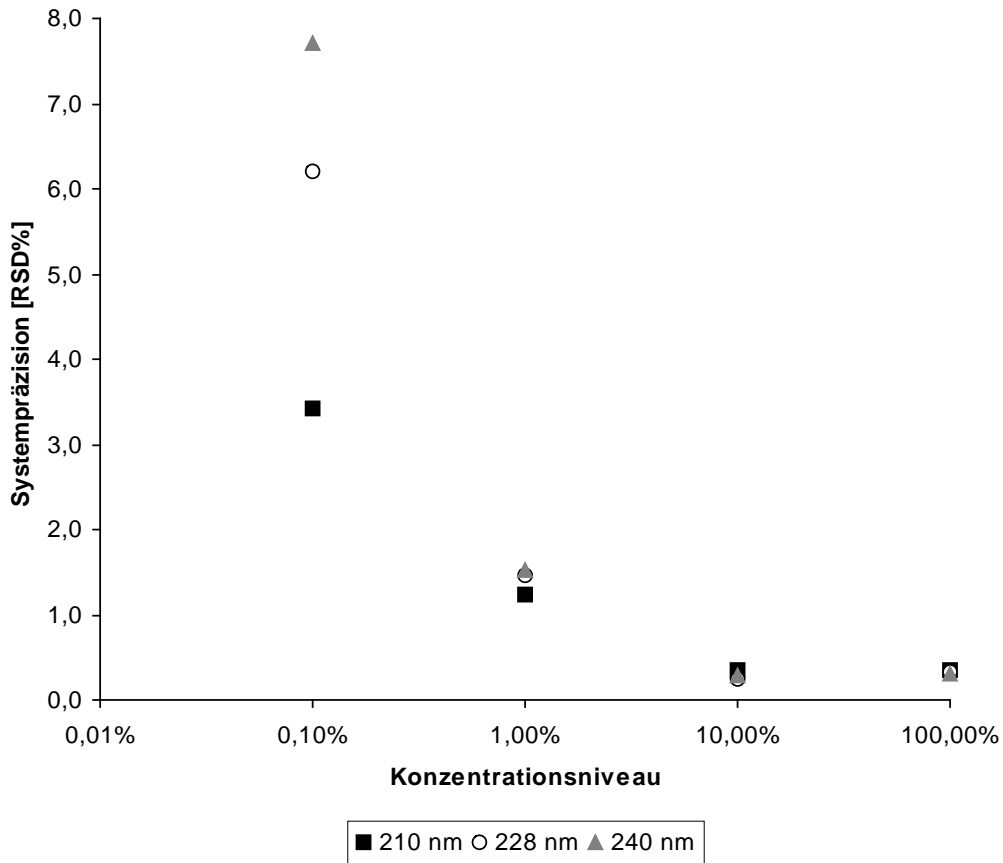


Abbildung 3.2-12: Darstellung der Systempräzision von (1) bei verschiedenen Wellenlängen zeigt ab einer Konzentration von 1,0 % einen dann gegen kleinere Konzentrationen stetig wachsenden Effekt auf die $\hat{\sigma}_{\text{sys}}$.

Bei Verwendung der ungünstigeren Wellenlänge sinkt die absolute Peakfläche und folglich das Signal-Rausch Verhältnis, was einen Anstieg des Integrationsfehlers zur Folge hat, der sich in einer schlechteren $\hat{\sigma}_{\text{sys}}$ niederschlägt. Die UV-Absorptionen 210, 228 und 240 nm stehen zueinander im Verhältnis 2,2 : 1,4 : 1,0. Das Signal-Rausch Verhältnis bei 210 nm ist für gleiche Konzentrationen ca. doppelt so groß. Vergleicht man eine Konzentration bei 240 nm gegenüber der halben Konzentration bei 210 nm, so erhält man das gleiche Signal-Rausch Verhältnis. Bei geringen Konzentrationen hängt die $\hat{\sigma}_{\text{sys}}$ nur noch vom Signal-Rausch Verhältnis ab, unabhängig davon, welche

Wellenlänge oder Konzentration dabei betrachtet wird. Dies zeigt ein weiteres Mal den hohen Stellenwert von korrekter Integration.

3.2.3 Relative Zusammensetzung der Wiederhol-Standardabweichung

Die Hauptvarianzkomponenten der $\hat{\sigma}_r$, die als gesamte Variation innerhalb eines Tages aufgefasst werden soll, sind die Probenaufbereitung und die Systempräzision, bestehend aus Injektions- und der Integrationsfehler, die in Tabelle 3.2-4 als fettgedruckte Terme markiert sind. Betraglich spielt die Probenaufbereitung eine große Rolle, aber während der Messreihen bleibt die relative Zusammensetzung der Probenaufbereitung im Wesentlichen unverändert und daher wird die Probenaufbereitung im Rahmen dieser Arbeit als Ganzes betrachtet. Der Beitrag der Variation aufgrund von Schwankungen in der Flussrate zur Systempräzision wird als zusätzliche Komponente vernachlässigt und nimmt üblicherweise ca. 10 % ein [54].

Wiederhol-Standardabweichung			
Probenaufbereitung		Systempräzision	
Einwaagefehler	Verdünnungsfehler	Injektionsfehler	Integrationsfehler
Inhomogenitäten [#]	Extraktion des Analyten [#]		

Tabelle 3.2-4: Variationskomponenten der $\hat{\sigma}_r$.

[#]: nur relevant für authentische Analyse von Tabletten, daher hier nicht betrachtet.

Alle Werte für die $\hat{\sigma}_r$ und die $\hat{\sigma}_{sys}$ als relative Standardabweichung sind bereits in Tabelle 3.2-1 aufgeführt worden. Für den Konzentrationsbereich von 100 bis 10 % konnten die $\hat{\sigma}_r$ und die $\hat{\sigma}_{sys}$ als konstant angesehen werden, so dass die Bestimmungen der Präzisionen auf den einzelnen Konzentrationsleveln als wiederholte Schätzung der Präzision aus derselben Grundgesamtheit aufgefasst

werden können. Als gepoolte RSD%-Werte für die $\hat{\sigma}_r$ und die $\hat{\sigma}_{\text{sys}}$ ergaben sich 0,70 bzw. 0,31 % (s. 3.2.1.1).

Da jeder Einzelwert aus einer eigenen unabhängigen Probenaufbereitung und einer anschließenden Messung entstanden ist, ergibt sich nach dem Gaußschen Fehlerfortpflanzungsgesetz die Gesamtvarianz nach folgender Gleichung 3.2-3 [25]:

$$\hat{\sigma}_{\text{ges}}^2 = \sum \hat{\sigma}_i^2 \quad (3.2-3)$$

Die Gesamtvarianz, die sich aus unabhängigen Varianzkomponenten zusammensetzt, ist gleich der Summe der gewichteten Einzelvarianzen. Für die in dieser Arbeit betrachtete gesamte Varianz, also dem Quadrat der $\hat{\sigma}_r$, ergibt sich 0,50. Die durch die Systempräzision bedingte Varianzkomponente beträgt 0,10 (= 0,31²). Die Differenz der Varianzkomponenten Wiederhol-Standardabweichung und Systempräzision ergibt 0,40 (= 0,50 - 0,10) und ist durch die Probenaufbereitung bedingt. Die Probenaufbereitung hat als RSD% damit den Wert 0,63 %.

Die Probenaufbereitung ist in diesen höheren Bereichen der Probenkonzentration die Hauptfehlerquelle. Der Beitrag der Probenaufbereitung kann als konstant aufgefasst werden. Während der Herstellung der drei Konzentrationsbereiche (s. Tab. 6.7-2) wird nur die anfängliche Einwaage reduziert; darüber hinaus unterscheiden sich die Spalten durch jeweils einen weiteren Verdünnungsschritt.

Die $\hat{\sigma}_{\text{sys}}$ verhält sich in diesem Bereich ebenfalls nahezu konstant, während die Probenkonzentration und damit die zu integrierende Fläche auf ein Zehntel sinkt (s. Abb. 3.2-3). Folglich muss der Integrationsfehler in diesem Bereich von untergeordneter Bedeutung sein. Ein Integrationsfehler relevanter Größe wäre von 100 % aus in Richtung 10 % merklich gewachsen, da die zu integrierenden Flächeneinheiten in diesem Intervall auf ein Zehntel sinken, und hätte zu einer Zunahme der $\hat{\sigma}_{\text{sys}}$ geführt.

Dieser Sachverhalt wird in Abbildung 3.2-13 durch die logarithmische Darstellung der $\hat{\sigma}_{\text{sys}}$ gegen die Konzentration veranschaulicht.

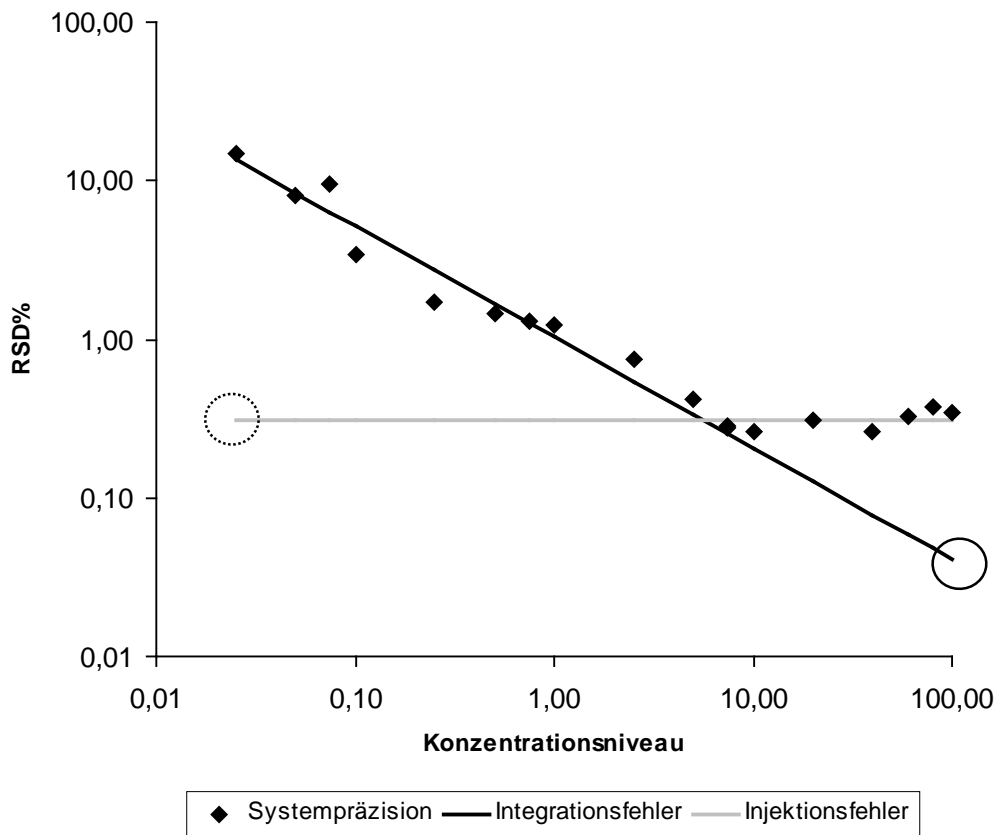


Abbildung 3.2-13: Doppelt logarithmische Darstellung der Systempräzision bestehend aus Injektionsfehler (dunkelgraue Diagonale) und Integrationsfehler (hellgraue Waagerechte) gegen die Konzentration.

Ausgehend von kleinen Konzentrationen zu größeren hin (bis 1 %) verläuft die Systempräzision wie eine Exponentialfunktion und lässt sich in einer doppelt logarithmischen Darstellung durch eine Diagonale beschreiben. Da in diesem Bereich die Systempräzision vom Integrationsfehler maßgeblich bestimmt wird, repräsentiert die Diagonale vornehmlich den Integrationsfehler, dessen RSD%-Wert im weiteren Verlauf gegen höhere Konzentrationen exponentiell absinkt. Extrapoliert auf den höchsten Konzentrationsbereich um 100 % (durchgezogener Kreis) ist zu erkennen, dass der Integrationsfehler dort praktisch keine Rolle mehr spielt.

Die Systempräzision ist fast vollständig durch den Injektionsfehler bedingt, der als konstante Fehlerquelle als Horizontale in Abbildung 3.2-13 eingetragen ist. Daher

verläuft die Systempräzision zwischen 100 % und 10 % ebenfalls konstant. Der prozentuale Beitrag des Injektionsfehlers wird bei kleineren Konzentrationen monoton geringer und wird bei sehr kleinen Konzentrationen (gepunkteter Kreis) schließlich vernachlässigbar klein.

Als wesentliche Varianzkomponente der $\hat{\sigma}_{\text{sys}}$ im hohen Konzentrationsbereich verbleibt der Injektionsfehler. Der für den Bereich von 100 bis 10 % berechnete Wert von 0,31 % dient daher als Schätzung des Injektionsfehlers über den gesamten Messbereich. Für Konzentrationen unter 0,5 % steigt die $\hat{\sigma}_r$ drastisch an. Der Beitrag der Probenaufbereitung zur $\hat{\sigma}_r$ wird daher gegen kleine Konzentrationen stetig geringer. Ein konstanter Fehler von 0,63 % durch Probenaufbereitung trägt gemäß Gleichung 3.2-1 bei einer Gesamtvariation von über 2 % (ab einer Probenkonzentration unterhalb von 0,5 %) nur noch weniger als 10 % bei (s. Abb. 3.2-14).

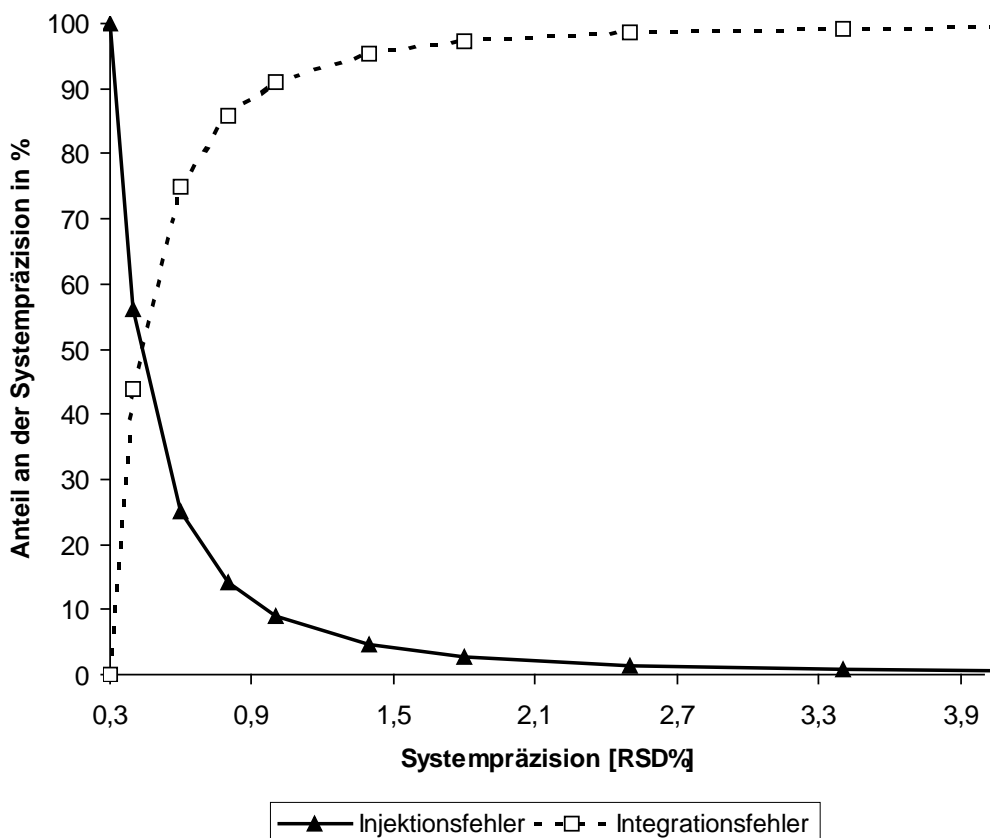


Abbildung 3.2-14: Die Systempräzision steigt dadurch an, dass der Integrationsfehler, der anfänglich vernachlässigbar klein ist, den Injektionsfehler einholt (zwischen 0,3 und 0,9 %) und im weiteren Verlauf anwächst, bis er der dominierende Fehlerbeitrag ist (spätestens ab 1 %).

Gegenüber dem konstanten Injektionsfehler wird der Integrationsfehler in diesen Konzentrationsbereichen ($< 0,5\%$) die dominante Variationskomponente, was in Reihe B bestätigt wird. Durch die Erhöhung des Injektionsvolumens um den Faktor 10 und die Konstanthaltung aller anderen Faktoren wurde der Verlauf der Präzision der Reihe B, verglichen mit Reihe A, um eben diesen Faktor 10 zu kleineren Konzentrationen verschoben (s. 3.2.1.2).

Nachdem in diesem Abschnitt die beiden konstanten Fehlerkomponenten Probenaufbereitung und Injektionsfehler in ihrer Höhe bemaßt werden konnten, zeigt Abbildung 3.2-15 den prozentualen Beitrag dieser Varianzkomponenten und zusätzlich des Integrationsfehlers zur gesamten Wiederhol-Standardabweichung in Abhängigkeit der Höhe der Wiederhol-Standardabweichung.

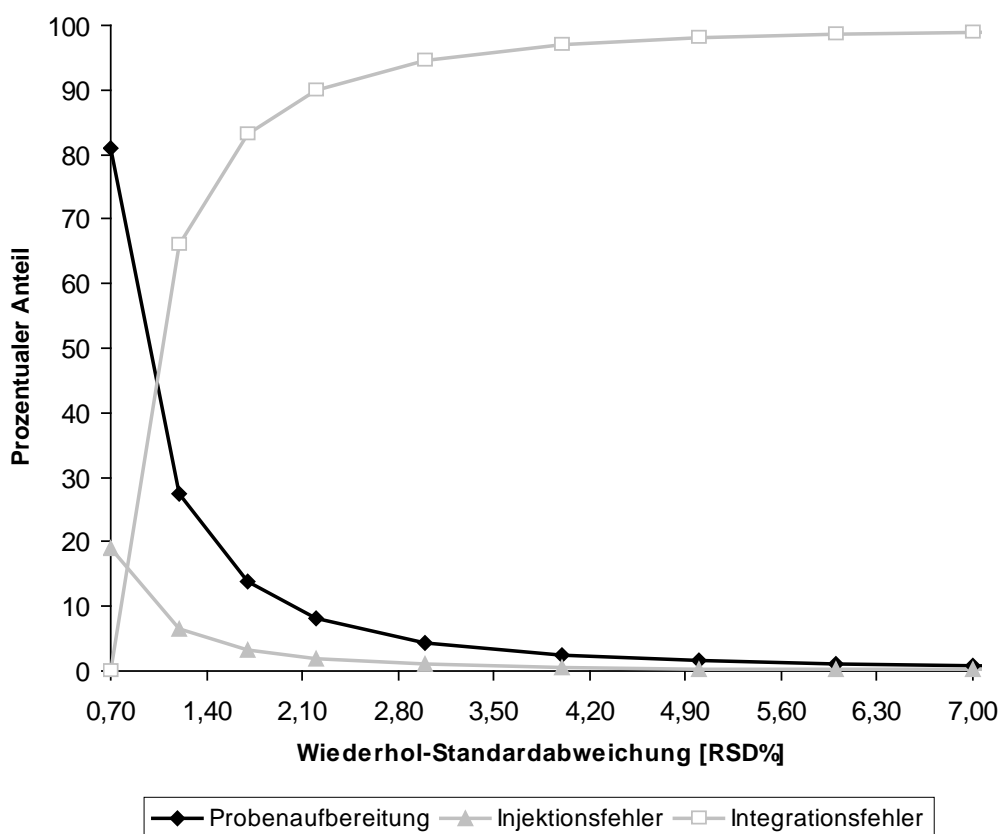


Abbildung 3.2-15: Anteil der einzelnen Variationskomponenten an der Wiederhol-Standardabweichung in Prozent.

Die Wiederhol-Standardabweichung in den verschiedenen Bereichen setzt sich sehr unterschiedlich zusammen. Die Gesamtpräzision im Bereich von 0,70 bis 1,0 % kann im wesentlichen nur durch eine Verbesserung der Probenaufbereitung verringert werden. Andererseits brächte eine verbesserte Probenaufbereitung deutlich oberhalb von 1,0 % im Grunde keinen Vorteil. In diesen Bereichen kann die Gesamtpräzision nur durch eine Erhöhung der Empfindlichkeit und die damit verbundene Verringerung des Integrationsfehlers verbessert werden (wie in Reihe B geschehen).

Abbildung 3.2-16 gibt zusammenfassend die relative Zusammensetzung der Wiederhol-Standardabweichung aus Probenaufbereitung und Systempräzision der eingesetzten HPLC-Methode schematisch wieder.

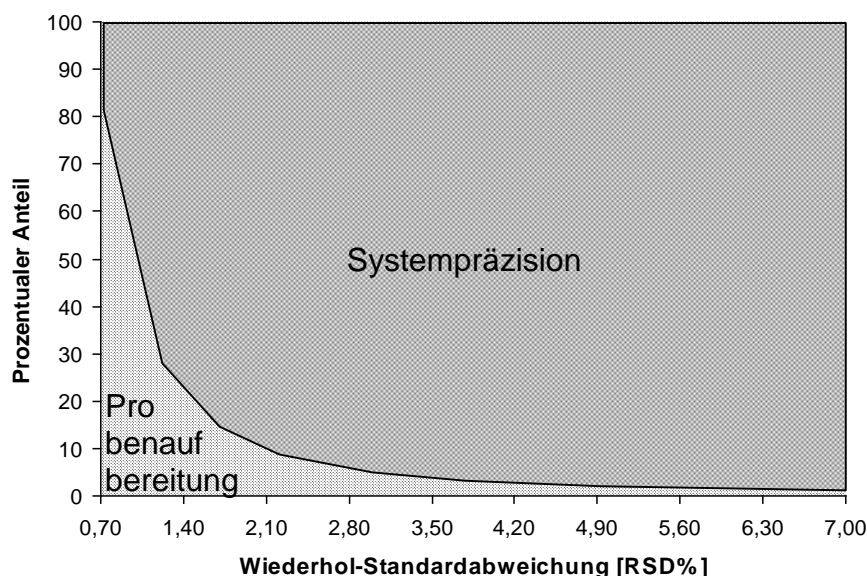


Abbildung 3.2-16: Übersicht über die Zusammensetzung der Wiederhol-Standardabweichung aus Probenaufbereitung und Systempräzision.

3.2.4 Abschätzung der Mehrtages-Standardabweichung

Zur Ermittlung der Mehrtages-Standardabweichung $\hat{\sigma}_{\text{int}}$ wurden die an den verschiedenen Tagen erhaltenen Messwerte zur Wiederfindung aus Reihe A herangezogen. Zwischen den einzelnen Serien ergeben sich durch die Herstellung verschiedener Standards, der mobilen Phase und des Lösungsmittels zusätzliche analytische Varianzkomponenten. Auch der Unterschied der jeweiligen Tagestemperatur fließt mit ein.

Als Konzentrationsbereich wurde das Intervall von 100 bis 10 % herangezogen, da sich die einzelnen Serien, wie in 3.2.1.1 erläutert, als Schätzungen aus der gleichen Grundgesamtheit auffassen lassen. Die betreffenden Einzelwerte sind in der nachfolgenden Abbildung 3.2-17 dargestellt:

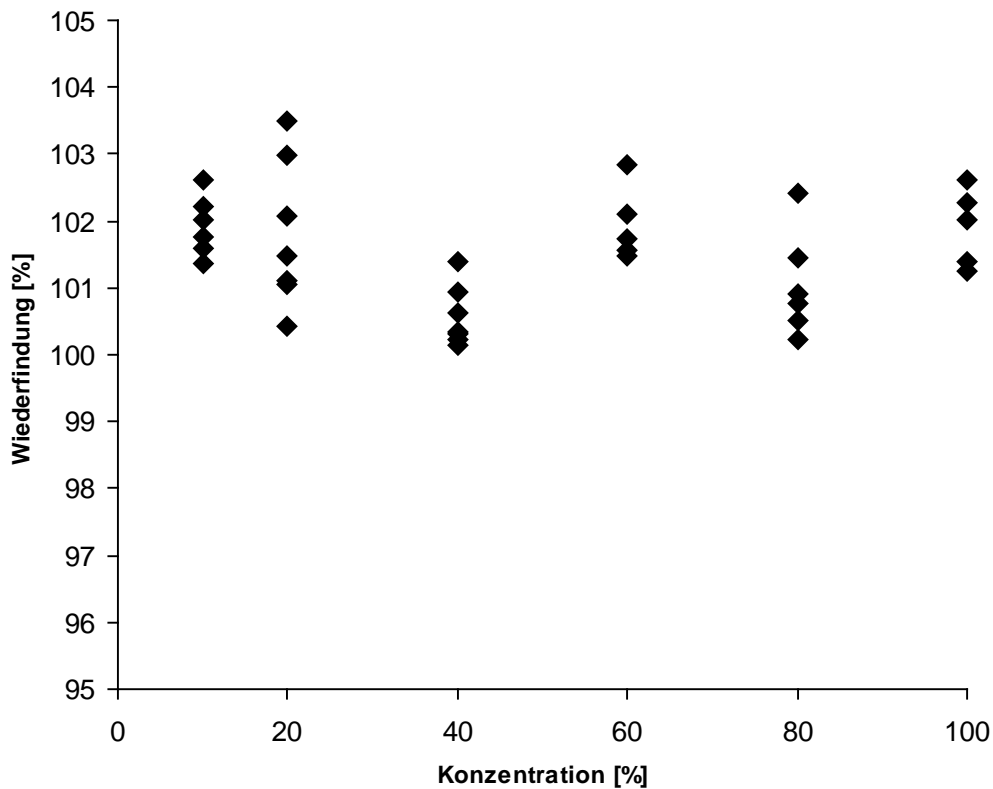


Abbildung 3.2-17: Übersicht über die zur Abschätzung der $\hat{\sigma}_{\text{int}}$ eingesetzten Einzelwerte aus Reihe A.

Die Mehrtages-Standardabweichung setzt sich aus den Varianzkomponenten Variation innerhalb der Serien $\hat{\sigma}_r$ und Variation zwischen den Serien $\hat{\sigma}_{zw}$ zusammen und lässt sich im Rahmen einer hierarchischen Varianzanalyse nach DIN ISO 5725-2 berechnen [55, 56]:

$$\hat{\sigma}_{\text{int}}^2 = \hat{\sigma}_r^2 + \hat{\sigma}_{zw}^2 \quad (3.2-4)$$

Die Wiederhol-Standardabweichung $\hat{\sigma}_r$ wurde bereits in Abschnitt 3.2.1.1 anhand von Gleichung 3.2-1 berechnet und ergab einen Wert von 0,70 %. Die zusätzliche Variation zwischen den Serien $\hat{\sigma}_{zw}$ berechnet sich nach:

$$\hat{\sigma}_{zw}^2 = \left(\frac{\sum_{i=1}^6 (n_i \cdot \bar{x}_i^2) \sum_{i=1}^6 n_i - \sum_{i=1}^6 (n_i \cdot \bar{x}_i)}{(6-1) \cdot \sum_{i=1}^6 n_i} - \hat{\sigma}_r^2 \right) \cdot \frac{(6-1) \cdot \sum_{i=1}^6 n_i}{\left(\sum_{i=1}^6 n_i \right)^2 - \sum_{i=1}^6 (n_i^2)} \quad (3.2-5)$$

Für die $\hat{\sigma}_{zw}$ wurde ein RSD%-Wert von 0,52 % erhalten und die $\hat{\sigma}_{\text{int}}$ berechnet sich schließlich zu 0,88 %. Damit setzt sich die Gesamtvariation der Mehrtages-Standardabweichung zu 65 % aus der Wiederhol-Standardabweichung und zu 35 % aus der Variation zwischen den Serien zusammen. Es ist anzunehmen, dass bei Miteinbeziehung eines größeren Zeitraumes und mehrerer Analytiker sowohl der Gesamtbetrag der $\hat{\sigma}_{\text{int}}$ als auch der prozentuale Beitrag der $\hat{\sigma}_{zw}$ steigen würden.

3.3 Ergebnisse zur Wiederfindung

3.3.1 Ergebnisse der Reihe A

Die Wiederfindung war über den ganzen Konzentrationsbereich systematisch erhöht und lag bei ca. 101,5 %. Diese Erhöhung ist statistisch hochsignifikant. So ergibt die Berechnung des einseitigen t-Testes für die Konzentration 120 % (gegen den Sollwert 100 %) bereits bei relativ kleiner Stichprobengröße einen signifikanten Unterschied zum Niveau 1 %. Die Wiederfindungsraten der einzelnen Konzentrationen sind in Tabelle 3.3-1 dargestellt. Die systematische Erhöhung gegenüber 100 % war verhältnismäßig konstant, berücksichtigt man die Zunahme der Variabilität gegen kleine Konzentrationen. Wie die Präzisionen ist auch die Wiederfindung für 7,5 % zwei mal überlappend bestimmt worden.

Konzentrations-niveau [%]	Wiederfindung [%]
120	101,43
100	101,91
80	101,05
60	101,94
40	101,57
20	101,80
10	101,93
7,5	102,28
7,5	104,05
5,0	104,77
2,5	104,62
1,0	101,51
0,75	102,30
0,50	102,21
0,25	106,93
0,10	102,75
0,075	101,32
0,050	93,28
0,025	92,13

Tabelle 3.3-1: Wiederfindungsraten der einzelnen Konzentrationen aus Reihe A

Oberhalb von 7,5 % sind die Proben auf einen Standard von 100 % und unterhalb von 7,5 % auf einen Standard von 1,0 % bezogen worden. Der Wechsel der

Konzentration des Standards von 100 zu 1,0 % zeigte keinen erheblichen Effekt auf die Wiederfindung. Abbildung 3.3-1 zeigt die Werte der Tabelle 3.3-1 in logarithmischer Darstellung.

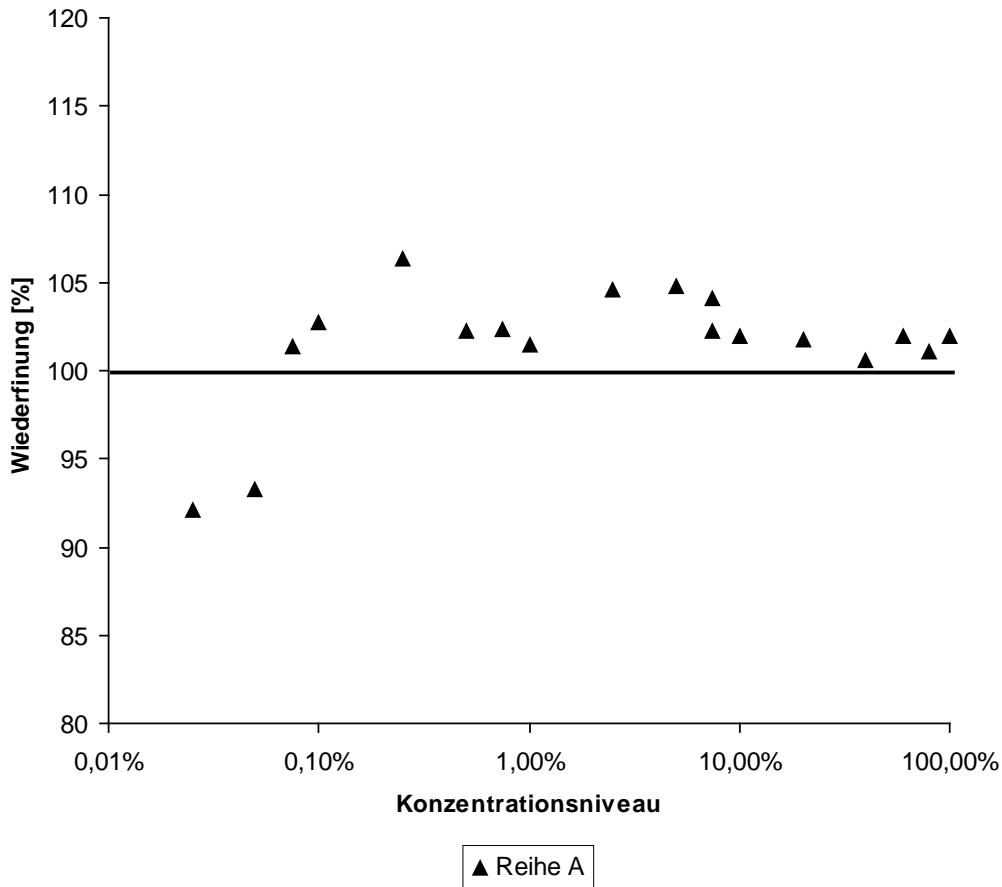


Abbildung 3.3-1: Verlauf der Wiederfindungsrate gegen die Konzentration in Reihe A. Unterhalb von 0,075 % fällt die Wiederfindung deutlich ab.

Der Grund für die erhöhte Wiederfindung konnte erst nach einigen Testungen gefunden werden. So wurden u.a. die Matrixbestandteile einzeln daraufhin untersucht. Dazu wurden Proben vermessen, denen neben (1) jeweils einzeln die Matrixkomponente in der entsprechenden Einwaage zugesetzt wurde (s. 6.7.6). Alle Matrixbestandteile führten zur gleichen, erhöhten Wiederfindung, selbst der Zusatz von 0,4 mg $\text{Fe}_2\text{O}_3 \cdot \text{H}_2\text{O}$. Folglich konnte keine dieser Komponenten als Ursache in Frage kommen.

Schließlich konnte ein Probenaufbereitungsschritt, der nur bei den Proben, nicht aber bei den Standards durchgeführt worden war, als Grund identifiziert werden.

Die Zentrifugation der matrixhaltigen Proben während der Probenaufbereitung (s. 6.7.6), wurde einer genaueren Prüfung unterworfen. Denkbar war eine Lösungsmittelverdunstung während des Zentrifugierens und die damit verbundene Aufkonzentrierung der Probe. Dies würde auch erklären, warum bei deutlich unterschiedlichen Konzentrationen stets die gleiche Abweichung gefunden wurde.

Um dies zu untersuchen, wurden jeweils vier Standards und Proben der Konzentration 100 % aufbereitet. Anschließend wurden sowohl die Proben als auch die an sich klar gelösten Standards zentrifugiert. Die Differenz zwischen beiden gefundenen Peakflächen der Proben und Standards wurde geringer, unabhängig davon, ob in offenen oder geschlossenen Vials zentrifugiert wurde. Aus dieser Serie ergab sich eine Wiederfindung von 100,5 %. Damit konnte der Hauptteil der Abweichung von 100 % identifiziert werden. Die restliche Abweichung lässt sich mit der Volumenverdrängung des Lösungsmittels durch die Tablettenmatrix bei den Proben erklären, was zur Aufkonzentrierung führte.

Weiterhin zeigt Abbildung 3.3-1 für die beiden kleinsten Konzentrationen einen Abfall der Wiederfindung auf unter 95 %. Ein Grund dafür liegt in der unterschiedlichen Integration des im Vergleich zum Rauschen kleinen 0,025 oder 0,050%-Peaks gegenüber dem 1,0%-Peak des Standards. Ein ausgeprägter Effekt der Integrationsparameter und Filter auf die Peakfläche für kleine Konzentrationen wird in einer Studie zur Präzision bei chiralen Trennungen in der Kapillarelektrophorese beschrieben [57]. Dort gab es bei hohen Konzentrationen ebenfalls keinen Einfluss der Parameter.

Ein verzerrtes Setzen der Basislinie aufgrund unzureichender digitaler Auflösung hat dann einen geringen Effekt, wenn Proben- und Standardpeak von gleicher Größenordnung sind. Je größer der Unterschied, desto stärker ist der relative Fehler durch verzerrtes Setzen einer Basislinie.

Um den Einfluss von Glättung (Smoothing) bei unzureichender digitaler Auflösung zu ermitteln, wurden die beiden letzten Konzentrationen mit Hilfe der Glättungsfunktion der HPLC-Software (sogenanntes 5pt Smoothing) integriert. Tabelle 3.3-2 zeigt die Wiederfindungsraten der beiden kleinsten Konzentrationen.

5pt Smoothing	ohne	mit
Konzentrations-niveau [%]	Wiederfindung [%]	Wiederfindung [%]
0,050	93,28	96,69
0,025	92,13	100,28

Tabelle 3.3-2: Einfluss der Glättung auf die Wiederfindungsrate.

Unter Verwendung der Glättungsfunktion näherte sich die Wiederfindung an das 100 % Niveau an. Der verzerrende Effekt unzureichender digitaler Auflösung im Grenzbereich und der Effekt der Glättung auf die Basislinie sind in Abbildung 3.3-2 dargestellt.

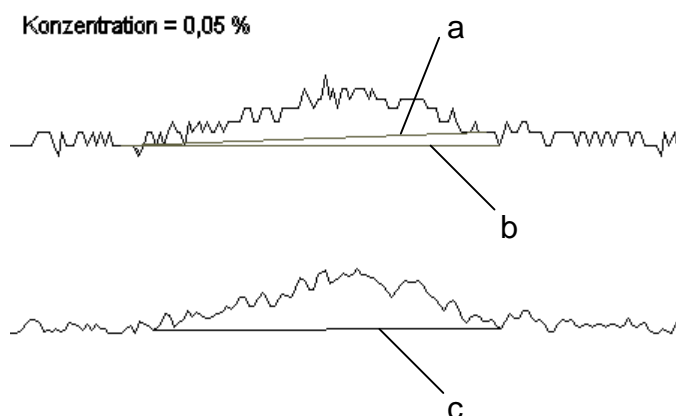


Abbildung 3.3-2: Die Abbildung zeigt einen Ausschnitt aus einem Chromatogramm der Konzentration 0,050 % aus Reihe A im normal mode. Der obere Teil ist nicht und der untere Teil mit zusätzlichem 5pt Smoothing bearbeitet worden.

Im oberen Ausschnitt sind zwei unterschiedliche, aber nachvollziehbar gesetzte Basislinien (a und b) für einen kleinen Peak im „normal mode“ (s. 6.4.1) gezeigt. Der relative Unterschied zwischen den erhaltenen Peakflächen ist groß. Die Basislinie in a führt zu einer Unterschätzung der tatsächlichen Fläche, wie es in

Reihe A bei 0,050 und 0,025 % der Fall war. Die wahre Basislinie wäre zwischen diesen beiden Optionen zu wählen (Linie c). Die 5pt Glättung der HPLC-Software erzeugt in der digitalen Auflösung neue Zwischenpunkte, die eine geeignetere Setzung der Basislinie ermöglichen.

3.3.2 Ergebnisse der Reihe C

In Reihe C sind die Wiederfindungsraten für die Verunreinigungen (3-6) bestimmt worden. Der Konzentrationsbereich erstreckte sich von 1,0 bis 0,025 %. Die einzelnen Wiederfindungsraten im Verlauf von 1,0 bis 0,025 % sind in Abbildung 3.3-3 graphisch dargestellt.

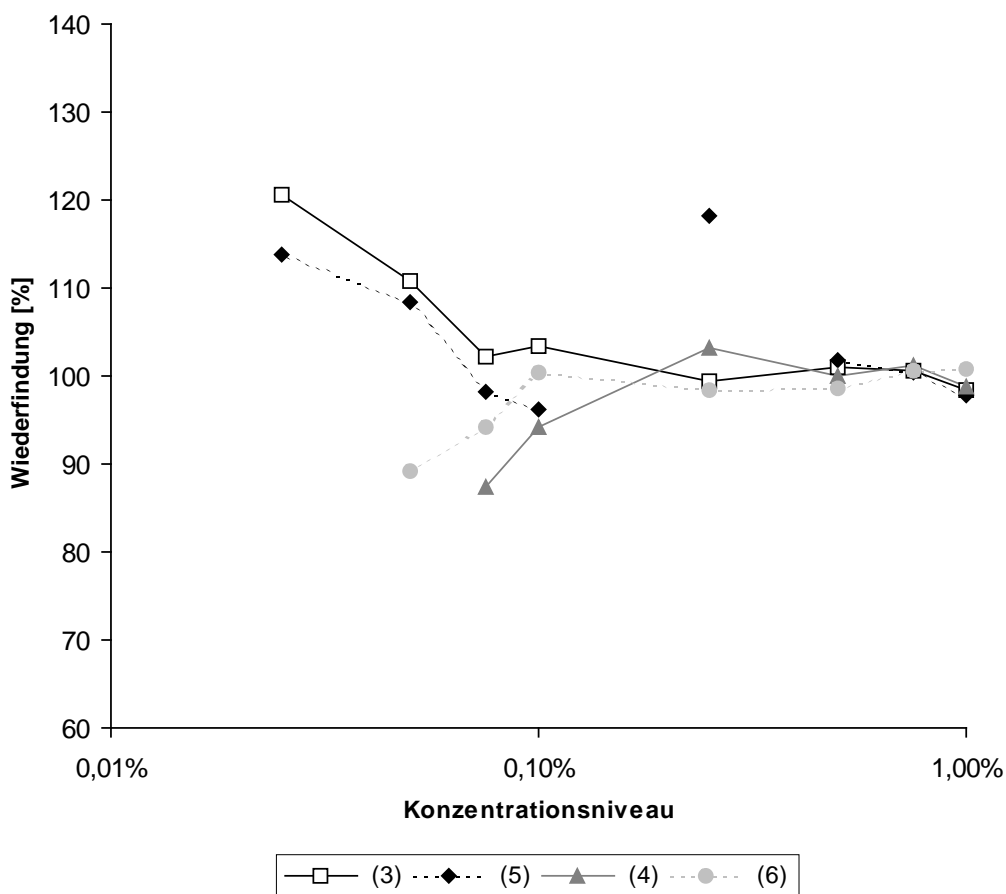


Abbildung 3.3-3: Die Wiederfindungsraten von (3) bis (6) aus der Reihe C. Die stark erhöhte Wiederfindung von (5) bei 0,25 % ist aufgrund eines Fehlers in der Probenaufbereitung entstanden.

Von (4) und (6) waren die letzten Konzentrationen nicht mehr bestimmbar. Die Wiederfindungsraten verlaufen bis 0,1 % konstant um 100 %. Danach entstehen sowohl Abweichungen nach oben (von (3) und (5), die im Vergleich zum Standard Glibenclamid früher eluieren) als auch nach unten ((4) und (6) eluieren nach Glibenclamid). Diese Abweichungen sind wie in Reihe A auf den verzerrenden Effekt der unzureichenden digitalen Auflösung zurückzuführen.

Um zu ermitteln, ob das Glätten im Bereich unter 0,1 % einen zu Reihe A vergleichbaren Effekt auf die Wiederfindung besitzt, wurden die kleinsten drei Konzentrationen 0,075; 0,050 und 0,025 % ebenfalls mit Hilfe der Glättungsfunktion der HPLC-Software integriert. Die so erhaltenen Wiederfindungen sind in Abbildung 3.3-4 neben den durch standardmäßige Integration erhaltenen dargestellt.

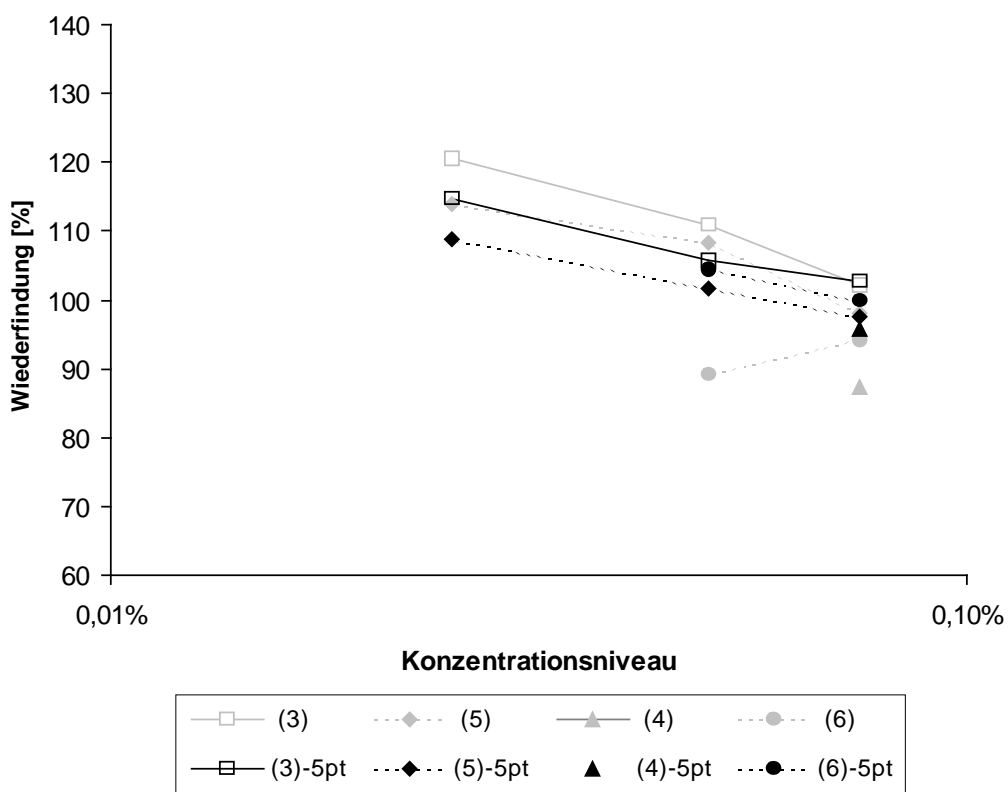


Abbildung 3.3-4: Die Wiederfindungsraten mit Hilfe des 5pt Smoothing liegen gegenüber denen bei standardmäßiger Integration näher an 100 %, weil mehr Möglichkeiten des Setzens einer Basislinie bestehen.

Neben dem Vergleich unterschiedlicher Peakflächen kommt hier noch zum Tragen, dass durch unterschiedliche Retentionszeiten unterschiedliche Peakhöhen resultieren. Bei Anwendung von Glättung kann ein prozentualer Anstieg der Wiederfindungsraten der Verunreinigungen (**3-6**) für kleine Konzentrationen angenommen werden. Diese resultieren möglicherweise aus schwierigen Integrationsvorgängen und dem verzerrten Setzen von Basislinien, was speziell bei den beiden kleinsten Konzentrationen der Fall ist. Selbst bei maximaler digitaler Auflösung und angemessener Glättung wird das Setzen der Basislinie gegen kleine Konzentrationen nahe der LOQ zunehmend problematischer. Zu beachten ist, dass Verschiebungen in der Basislinie auch Abweichungen von der Linearität verursachen können [57].

3.4 Vergleich von statistischem und empirischem Vertrauensbereich anhand der Mindestanforderungen der ICH-Richtlinie Q2B

Die ICH-Richtlinie Q2B [10] empfiehlt für die Bewertung der Richtigkeit im Rahmen der Validierung analytischer Methoden mindestens neun Bestimmungen der Wiederfindung. Diese Werte sollen sich über mindestens drei Konzentrationen des betrachteten Bereichs erstrecken (z.B. drei unabhängige Bestimmungen für drei Konzentrationen). Des Weiteren wird gefordert, die Richtigkeit samt statistischem Konfidenzintervall anzugeben.

Die Konzentrationen der Reihe A bestehen jeweils aus sechs Einzelwerten und gehen damit über das für eine einzelne Konzentration mindestens notwendige Maß von drei Bestimmungen hinaus (18 statt 9 Einzelwerte). Da die Wiederfindungswerte der Reihe A die oben beschriebene Verzerrung besaßen, wurden im Rahmen einer Nachmessung die Konzentrationen 80, 100 und 120 % erneut bestimmt.

Aus diesen 18 Werten lassen sich der Gesamtmittelwert, das 95%-Konfidenzintervall und das 95%-Prognoseintervall berechnen. Zusätzlich werden alle Möglichkeiten, aus diesem Datensatz durch Permutation 3 mal 3 Einzelwerte auszuwählen und davon den Mittelwert zu bilden ermittelt. Aus der nach Größe geordneten Aufreihung dieser Mittelwerte kann ein empirischer Vertrauensbereich gebildet werden.

Es sollte die Frage untersucht werden, wie sich die statistisch berechneten Intervalle zu dem durch Permutation gebildeten empirischen Vertrauensbereich verhalten.

Dazu wurden anhand eines Permutationsschemas alle Möglichkeiten, aus den drei Konzentrationen drei von sechs Werten zu ziehen, gebildet. Dies ergibt $20^3 = 8000$ Möglichkeiten.

Für jede dieser Möglichkeiten wurde die mittlere Wiederfindungsrate berechnet und anschließend der Größe nach geordnet (s. Abb. 3.4-1).

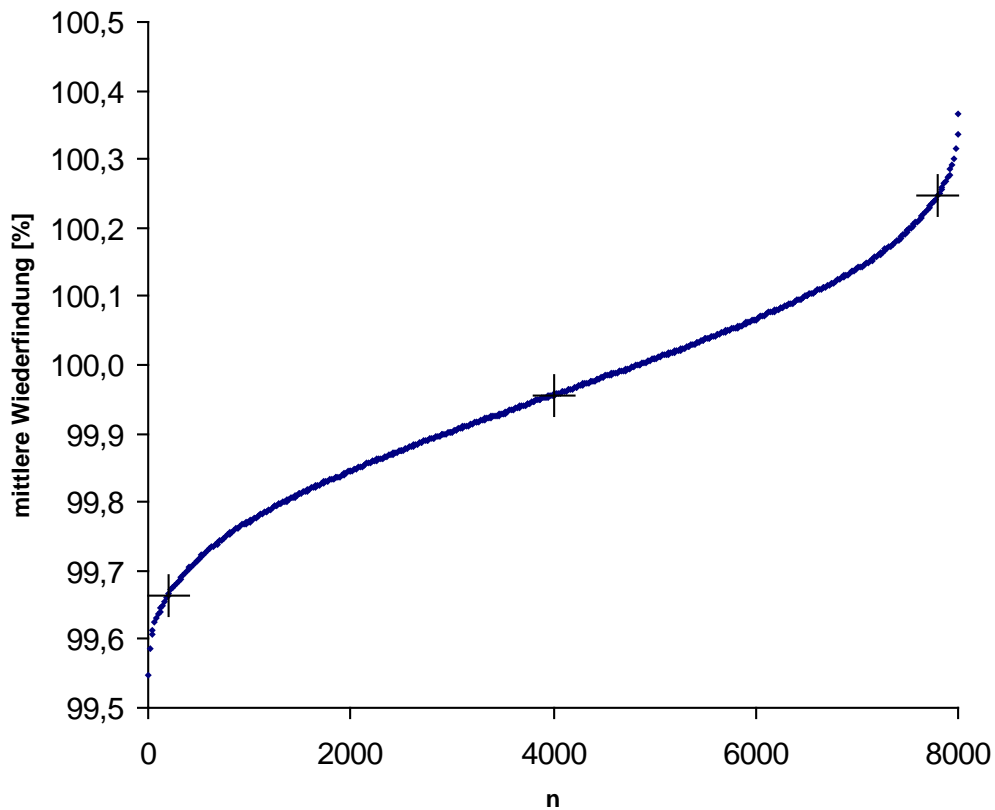


Abbildung 3.4-1: Der Größe nach geordnete mittlere Wiederfindungsraten. Als Kreuz markiert: untere Grenze (x_{200}), obere Grenze (x_{7800}) und zentraler Wert x_{4000} .

Aus dieser der Größe nach geordneten Reihe der mittleren Wiederfindungsrate lässt sich ein zweiseitiges Intervall bilden, in dem 95 % aller Werte liegen. Dazu werden jeweils oben und unten die Grenzwerte bestimmt, von denen aus 2,5 % der Werte außerhalb liegen. In dem hier abgebildeten Fall bilden die Werte $x_{200} = 99,66 \%$ und $x_{7800} = 100,25 \%$ damit die untere und obere Grenze des Intervalls um den zentralen Wert $x_{4000} = 99,96 \%$: $[99,66; 100,25]$.

Die Berechnung des Vertrauens- und des Vorhersagebereichs aus allen 18 Einzelwerten erfolgte anhand der Gleichungen 3.4-1 und 3.4-2 [58]:

$$cnf(\bar{x}) = \bar{x} \pm t_{0,975;17} \cdot \hat{\sigma} \sqrt{1/18} \quad (3.4-1)$$

$$prd(\bar{x}) = \bar{x} \pm t_{0,975;17} \cdot \hat{\sigma} \sqrt{(1/18) + (1/9)} \quad (3.4-2)$$

Für den Vertrauensbereich ergibt sich [99,63; 100,29] und für den Vorhersagebereich einer weiteren Stichprobe der Größe $n = 9$ das Intervall [99,38; 100,53].

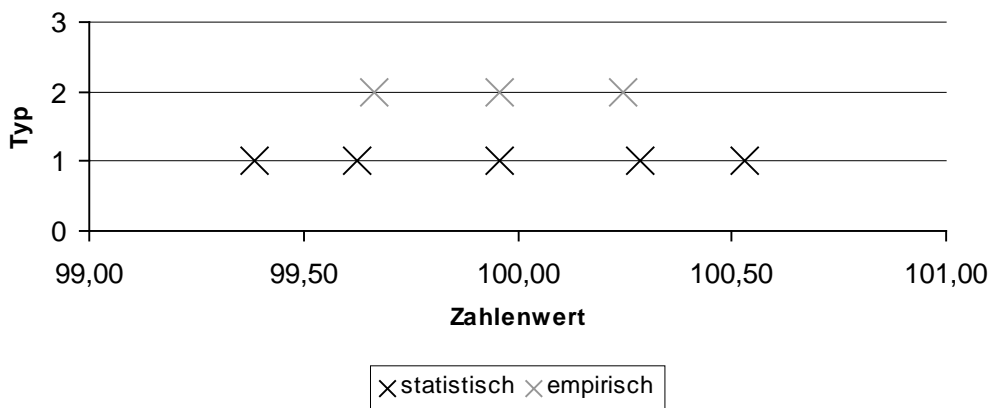


Abbildung 3.4-2: Darstellung der statistischen Parameter (Zeile 1) und der empirischen Grenzen (Zeile 2) auf einem Zahlenstrahl.

Interessanterweise zeigt der Vergleich beider Möglichkeiten in Abbildung 3.4-2, dass das empirische Intervall eng mit dem zweiseitigen 95%-Konfidenzintervall des ganzen Datensatzes übereinstimmt. Das Vorhersageintervall hingegen ist wesentlich breiter. Folglich liefert das empirische Intervall anhand der Permutationsdaten keine Information über die zukünftig zu erwartende Streuung. Die Übereinstimmung des empirischen Intervalls mit dem statistischen Vertrauensbereich belegt die korrekte Normalverteilungsannahme der empirischen HPLC-Daten des betrachteten Bereichs.

Dieser Zusammenhang konnte anhand von zusätzlich durchgeführten Simulationen der gleichen Prozedur mit künstlich normalverteilten Werten ähnlicher Zusammensetzung bestätigt werden.

4 Leistung von Äquivalenztests beim analytischen Methodentransfer

4.1 Notwendige Voraussetzungen

Damit die statistischen Tests, deren Leistungen in diesem Kapitel beurteilt werden, Gültigkeit besitzen, müssen gewisse Voraussetzungen bezüglich der Daten, an denen die Tests angewandt werden, erfüllt sein.

Eine notwendige Bedingung ist es, dass die Einzelwerte x_i , aus denen sich die Grundgesamtheit zusammensetzt, normalverteilt sind. Diese Voraussetzung kann hier problemlos eingehalten werden, da es sich um Simulationen handelt. In realen Testungen hingegen können Verletzungen dieser Normalverteilungsannahme vorkommen.

Des Weiteren setzen die hier eingesetzten Tests voraus, dass die Einzelwerte x_i , anhand derer die Prüfgröße für den statistischen Test berechnet wird, voneinander unabhängig sind. Anderenfalls können Berechnungen von abhängigen (korrelierten) Werten zu einer Verzerrung der Leistungsfähigkeit des eingesetzten Testes führen. So kann es z.B. vorkommen, dass ein statistischer Test im Rahmen einer Simulation die vorgegebene Irrtumswahrscheinlichkeit nicht einhält. Ob zwei analytische Messwerte voneinander unabhängig sind, hängt vom jeweiligen Versuchsdesign ab. Zwei Werte sind voneinander abhängig, wenn ihnen an irgendeiner Stelle im Versuchsdesign derselbe Arbeitsschritt (z.B. eine Verdünnung) zugrunde liegt. Ein Fehler während dieses Arbeitsschrittes wirkt sich auf die beiden Werte in der gleichen Weise aus. Trägt ein Arbeitsschritt allerdings sehr wenig zum Gesamtfehler bei, kann eine eventuelle Abhängigkeit nachfolgender Werte vernachlässigt werden.

4.2 Leistung beim Vergleich von Mittelwerten

4.2.1 Vereinfachung bezüglich der Varianzen

Prinzipiell ist eine sehr große Anzahl möglicher Konstellationen an unterschiedlichen Varianzen zwischen den Labors denkbar, die im Rahmen der

Simulationen getestet werden könnten. Für den Fall verschiedener Präzisionen hätten z.B. viele Fälle ähnlicher Präzisionen wie $\hat{\sigma}_1 = 0,50 \%$ und $\hat{\sigma}_2 = 0,55 \%$ oder $\hat{\sigma}_1 = 0,50 \%$ und $\hat{\sigma}_2 = 0,45 \%$ usw. unterschieden werden können. Als Vereinfachung für die Durchführung und zur Erhaltung der Übersichtlichkeit der Ergebnisse der Simulationen sind daher für beide Labors gleiche Gesamtvarianzen eingesetzt worden.

4.2.2 Designauswahl

Es wurden zwei verschiedene Versuchsanordnungen simuliert. Einerseits wurden die Simulationen anhand eines möglichst bekannten, einfachen und repräsentativen Designs mit üblichen Fehlerbeträgen und vorgegebener Akzeptanzgrenze vorgenommen, so dass an diesen Beispielen ein klarer Einblick in die auftretenden Annahmewahrscheinlichkeiten entsteht.

Andererseits wurde die Leistung des Äquivalenztestes in Abhängigkeit des %AkzG-Wertes abgebildet. Dieser Term beschreibt den prozentualen Anteil der betrachteten Präzision an der Akzeptanzgrenze und ist damit eine Verallgemeinerung, die von der Höhe der Akzeptanzgrenze unabhängig ist.

4.2.3 Leistungsfähigkeit des Äquivalenztestes ausgehend vom ISPE

Versuchsdesign

Zur Bewertung der Leistungsfähigkeit des Äquivalenztestes galt es, ein möglichst bekanntes, einfaches und repräsentatives Design zu wählen. Daher wurde ein Designvorschlag für Analysemethoden zur Bestimmung von Arzneiformen (active pharmaceutical ingredients) der ISPE (International Society of Pharmaceutical Engineering) herangezogen. Die ISPE ist ein internationaler Verband aus mehr als 20.000 Life Science Mitgliedern in 80 Ländern. Über Diskussionsforen, Weiterbildungsangebote und Veröffentlichungen bietet der Verband seinen Mitgliedern Möglichkeiten zur Entwicklung technischen Wissens, zum Austausch praktischer Erfahrungen und zur Zusammenarbeit in Bezug auf die Verbesserung industrieller Praktiken (www.ispe.org). In dem zum Thema Methodentransfer veröffentlichten Leitfaden „Technology Transfer“ wird folgendes empfohlen: “at

least two analysts at each laboratory should independently analyze three sample lots in triplicate; resulting in three distinct executions of the method” [19].

Dieser Designvorschlag lässt sich auf zwei verschiedene Arten interpretieren. Wird der Begriff „sample lot“ als einfache Stichprobe aufgefasst, so würden drei Stichproben gezogen und dreimal unabhängig voneinander analysiert. Die Variation zwischen den drei Stichproben kann in diesem Fall als sehr klein angesehen werden. Folglich entspricht diese Vorgehensweise einer neunmaligen Ausübung der Methode an einer Charge, so dass nur die Variation zwischen den Labors in Betracht gezogen wird.

Fasst man den Begriff „sample lot“ als Charge auf, dann geht man von drei unterschiedlichen Chargen aus, die jeweils dreimal unabhängig mit der Analysemethode bestimmt werden. Bei dieser Sichtweise kann getestet werden, ob die Methode auch dann noch in beiden Labors zu gleichen Ergebnissen führt, wenn sich das Verunreinigungsprofil der Chargen unterscheidet. Die Selektivität der Methode und ihre Robustheit werden also mitgetestet.

Anzumerken ist: Wenn alle Analytiker jeweils die selben drei Chargen analysieren, dann beeinflusst die zusätzliche Variation durch Unterschiede zwischen den Chargen nicht die Wahrscheinlichkeit, letztlich den Äquivalenztest zu akzeptieren. Unterscheidet sich die Variabilität in den Labors zwischen den Chargen, dann kann dies allerdings bei gleichen Mittelwerten ohne zusätzliche Messungen nicht entdeckt werden. So führen die drei Mittelwerte 97, 100 und 103 in Labor 1 und 99, 100 und 101 in Labor 2 zum gleichen Mittelwert 100.

Für die zuletzt genannte Interpretation sind weder die Verwendung eines gepaarten t-Testes noch die simultane Testung der drei Chargen auf einem fixen, globalen Signifikanzniveau geeignet. Diese Testungen setzen voraus, dass die Werte, die in diese Tests innerhalb eines Labors eingesetzt werden, voneinander unabhängig sind. Da drei unterschiedliche Chargen untersucht werden, werden die Einzelwerte zu einem Chargenmittelwert zusammengefasst. In dieser Fehlerstruktur ist der Analytiker keine unabhängige Fehlergröße mehr, denn für jede der drei Chargen werden dieselben Analytiker herangezogen. Je größer der Fehler durch den Analytiker ist, desto stärker ist die Abweichung von der eingesetzten Irrtumswahrscheinlichkeit.

Daher sind die Testdatensätze nach Möglichkeit 1 erstellt worden, die in Abbildung 4.2-1 dargestellt ist.

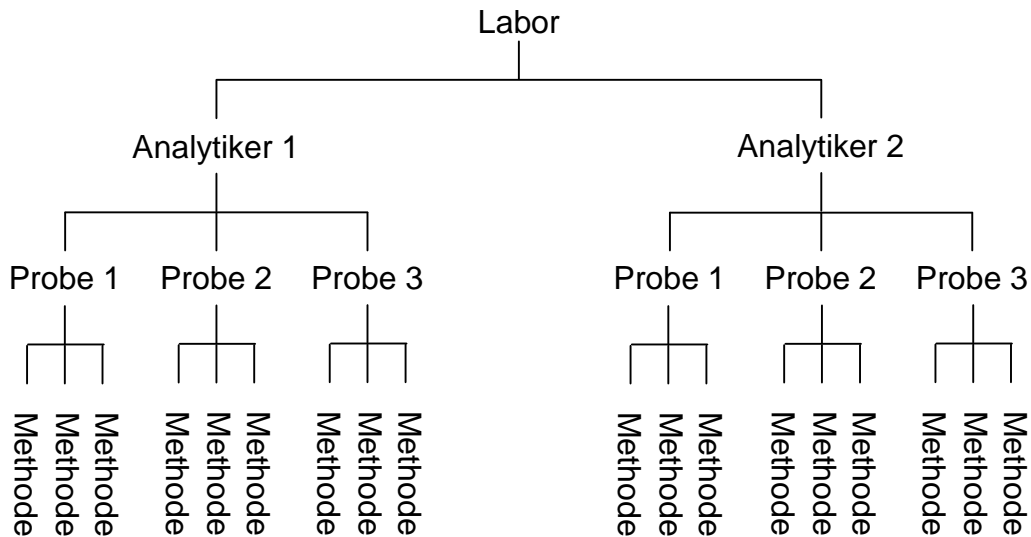
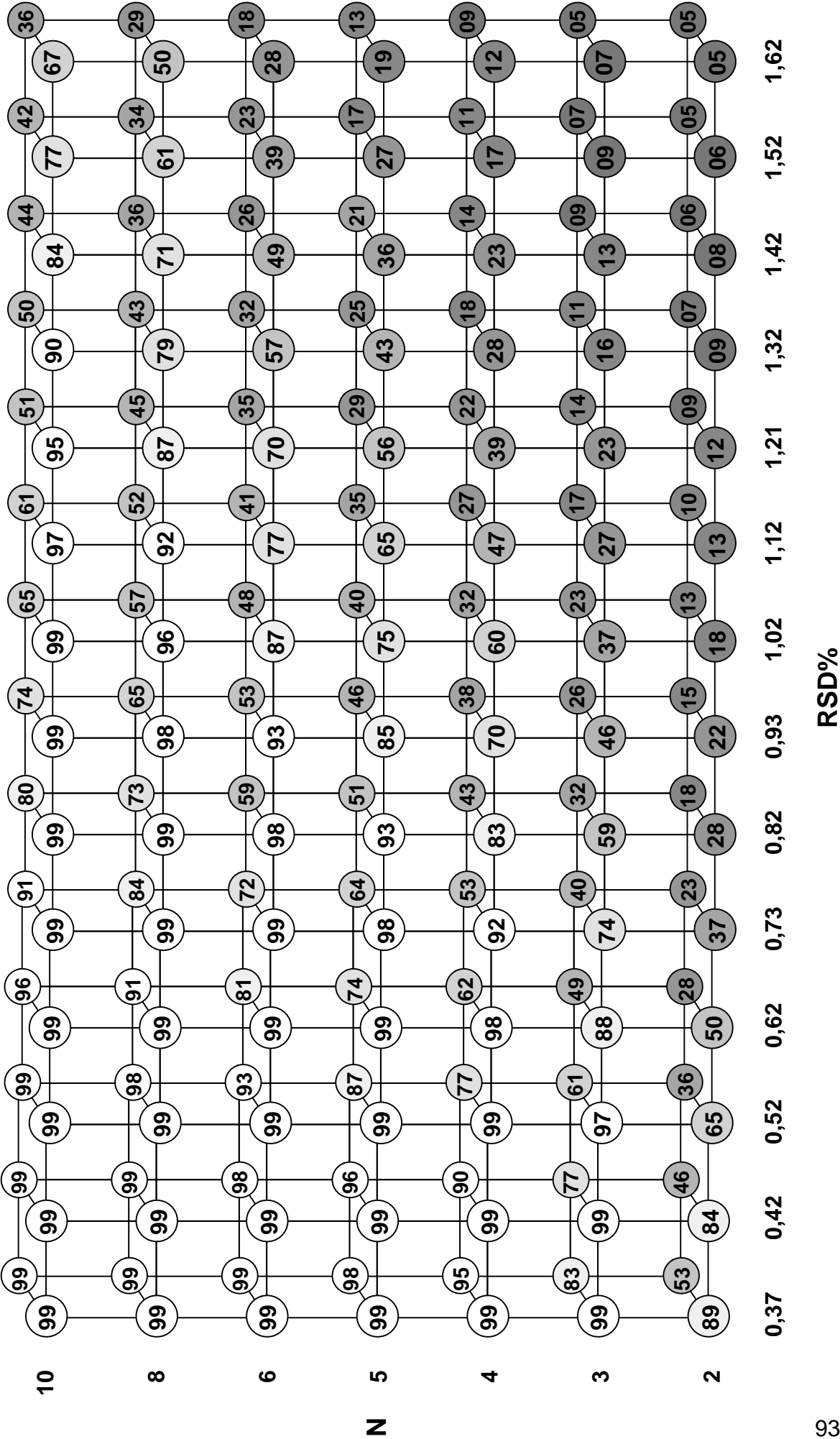


Abbildung 4.2-1: ISPE-Designvorschlag innerhalb eines Labors. Ein Analytiker bearbeitet 3 Stichproben, indem er an jeder unabhängig voneinander 3 mal die vollständige Methode anwendet (Probenaufbereitung und Messung).

4.2.4 Ergebnisse

Um einen ersten Einblick in die Leistungsfähigkeit des Äquivalenztestes zu bekommen, sind analytische Fehlerbeträge für die Simulation eingesetzt worden, die für eine Transfersituation üblich sind. Als Variationskomponenten der Methode wurden mit 0,60 % für $\hat{\sigma}_r$ und 0,30 % für $\hat{\sigma}_{sys}$ näherungsweise die Fehlerbeträge aus der in Kapitel 3 eingesetzten HPLC-Analytik (Konzentrationsbereich > 10 %) eingesetzt [59]. Für die Variationskomponente des Analytikers wurden zunächst 0,30 % eingesetzt.

Die Beträge dieser Komponenten führen nach Gl. 7.1-2 zu einer Gesamtvariation $\hat{\sigma}_{\bar{x}_{an}}$ von 0,37 %. Im weiteren Verlauf wurden unterschiedliche Beträge für die Variationskomponenten eingesetzt. Für verschiedene Gesamtvariationen der Labors und die Anzahl der eingesetzten Analytiker sind die entsprechenden Simulationen mit einem Äquivalenztest mit der Akzeptanzgrenze 2 % und $\alpha = 0,05$ durchgeführt worden. Die Ergebnisse sind als Annahmewahrscheinlichkeit (Power) in der nachfolgenden Abbildung 4.2-2 dargestellt und mit einer Unsicherheit von +/- 0,5 verbunden (s. 7.1.5).



(vorherige Seite) Abbildung 4.2-2: Die Werte in den Kreisen stehen für die prozentuale Wahrscheinlichkeit, den Methodentransfer anzunehmen (z.B. 88 = 88 %). Aus Gründen der Übersichtlichkeit sind nur zwei Ebenen mit unterschiedlichem Bias abgebildet: vordere Ebene Bias = 0 %, hintere Ebene Bias = 1 %. Auf der x-Achse ist der RSD%-Wert der Mittelwerte der Analytiker $\hat{\sigma}_{\bar{x}_{an}}$ bzw. Serien $\hat{\sigma}_{\bar{x}_s}$ der beteiligten Labors (fließen in den Äquivalenztest als Einzelwerte ein) aufgetragen (berechenbar nach Gl. 7.1-2). Die y-Achse zeigt die Anzahl N der an dem Transfer beteiligten Analytiker je Labor. Die Irrtumswahrscheinlichkeit beträgt 5 %.

Da der Bias in Abbildung 4.2-2 in beiden Ebenen kleiner als die Akzeptanzgrenze (2 %) ist, handelt es sich bei den Annahmewahrscheinlichkeiten gleichzeitig um die Power (= gerechtfertigte Annahme der Alternativhypothese). Der Wert für die Power von 89 % für eine Variation $\hat{\sigma}_{\bar{x}_{an}}$ von 0,37 % der beteiligten Labors mit zwei Analytikern für den Fall Bias = 0 zeigt, dass das ISPE-Design für diesem Bereich befriedigende Ergebnisse liefert.

Nähert sich der Bias der Akzeptanzgrenze von 2 %, fällt die Power stetig ab, bis sie bei einem Bias von 2 % (= Akzeptanzgrenze) genau den Wert 5 % erreicht (hier nicht abgebildet). An der Stelle $\hat{\sigma}_{\bar{x}_{an}} = 0,37 \%$ und einem Bias von 1 % beträgt die Power noch 53 %. Das zeigt, dass Situationen, in denen die wahre Abweichung zwischen den Labors im Verhältnis zur Akzeptanzgrenze groß ist, seltener durch den Äquivalenztest angenommen werden.

Für eine Transfersituation, bei der die $\hat{\sigma}_{\bar{x}_{an}} = 0,62 \%$ ist, fällt die Power auf 50 % ab. Daraus kann man schließen, dass das minimal vorgeschlagene Design nicht immer ausreichend ist. Würden hingegen drei Analytiker an dem Transfer beteiligt, so steigt die Annahmewahrscheinlichkeit auf 88 % (s. Abb. 4.2-2, zweite Zeile, vierte Spalte).

Ein Beitrag durch den Analytiker $\hat{\sigma}_{an}$ von 0,30 % zur $\hat{\sigma}_{\bar{x}_{an}}$ kann in bestimmten Situationen als zu gering angesehen werden. Als zusätzliche Fehlerquellen, die zu Erhöhung der Gesamtvariation führen, sind denkbar: z.B. Variation zwischen verschiedenen Tagen.

Um zu kennzeichnen, dass gegebenenfalls zusätzliche Fehlerquellen in Betracht gezogen worden sind, die die Variation zwischen den Serien der Analytiker vergrößern, wurde der Term $\hat{\sigma}_{an}$ durch $\hat{\sigma}_s$ ersetzt.

Da auch innerhalb einer Serie, die ein Analytiker bearbeitet, die Anzahl an Probenaufbereitungen n_{sp} oder Injektionen n_{sys} im Vergleich zum Vorschlag der ISPE variiert werden könnte, ist eine verallgemeinerte Formel zur Berechnung der Variation einer Serie $\hat{\sigma}_{\bar{x}_s}$ durch Gleichung 4.2-1 notwendig:

$$\hat{\sigma}_{\bar{x}_s} = \sqrt{\hat{\sigma}_s^2 + \frac{\hat{\sigma}_{sp}^2}{n_{sp}} + \frac{\hat{\sigma}_{sys}^2}{n_{sp} \cdot n_{sys}}} \quad (4.2-1)$$

Die $\hat{\sigma}_{sys}$ und die $\hat{\sigma}_{sp}$ besitzen nur einen kleinen Effekt auf die $\hat{\sigma}_{\bar{x}_s}$ (s. 7.1.6). Deshalb ist nur der Wert $\hat{\sigma}_s$ erhöht worden, um die Variation zwischen den Serien zu steigern. Die weiteren Werte der x-Achse ($\hat{\sigma}_{\bar{x}_s}$): 0,41; 0,52; 0,62 bis zu 1,64 in Abbildung 4.2-2 sind erzeugt worden, indem für den Term $\hat{\sigma}_s$ die Werte: 0,35; 0,47; 0,58 bis 1,60 eingesetzt wurden.

Gemäß dieser Verallgemeinerung zeigt Abbildung 4.2-2 den Effekt höherer Variationen der beteiligten Labors auf die Annahmewahrscheinlichkeit (durch Gl. 4.2-1 berechenbar).

Um eine Power von 80 %, die als ausreichend angesehen werden kann, nicht zu unterschreiten, muss (beginnend ab einer Variation von 0,41 %) für jede Erhöhung der Variation um 0,2 Prozentpunkte mindestens eine weitere Serie (inklusive Analytiker) in das Versuchdesign mit einbezogen werden.

Daran lässt sich erkennen, dass ein festes Design nicht grundsätzlich in der Lage ist, eine angemessene Power zu liefern. Vielmehr ist für eine spezielle Situation in Abhängigkeit von der erreichbaren Präzision eine individuelle Anzahl von Serien nötig. Bei einer Variation von 1,02 % und einer Anzahl von drei Serien liegt die Power bei nur noch 37 %. Ein höherer RSD%-Wert der Mittelwerte der Serien der Labors erfordert eine größere Anzahl von Serien, wenn man einen erfolgreichen Transfer mit hoher Wahrscheinlichkeit als solchen identifizieren möchte.

Aus Abbildung 4.2-2 sind folglich Annahmewahrscheinlichkeiten für eine Vielzahl unterschiedlicher Konstellationen zu entnehmen. Verschiedene Situationen, die nach Gleichung 4.2-1 zur gleichen $\hat{\sigma}_{\bar{x}_s}$ führen, besitzen die gleiche Power.

Mit der Kenntnis der eigenen Fehlerbeträge und der daraus resultierenden Variation der Serienmittelwerte ist es möglich, anhand der Abbildung 4.2-2 die entsprechende Power abzulesen.

4.2.5 Ergebnisse bei Verwendung verschiedener Chargen

Der Einfluss der Verwendung von drei unterschiedlichen Chargen auf die Leistungsfähigkeit bei Anwendung der simultanen Testung unter einem globalen Signifikanzniveau und des t-Testes für Paardifferenzen wurde untersucht.

Um das Ausmaß der abhängigen Fehlerquelle des Analytikers auf die simultane Testung zu ermitteln (s. 7.1.8), sind folgende Fehlerbeträge in einer Simulation eingesetzt worden: $\hat{\sigma}_{an} = 0,30 \%$, $\hat{\sigma}_{sp} = 0,60 \%$, $\hat{\sigma}_{sys} = 0,30 \%$ und ein Bias von 2,00 %. Dies führte an der Ablehngrenze zu einer Annahmewahrscheinlichkeit (α -Fehler) von 9 %. Folglich ist der Test nicht in der Lage, bei korrelierten Daten das Risiko, einen unzureichenden Transfer anzunehmen, auf 5 % zu kontrollieren.

Zusätzlich wurde der verzerrende Effekt korrelierter Daten auf die Anwendung des t-Testes für Paardifferenzen evaluiert (s. 7.1.9). Die Werte $\hat{\sigma}_{an} = 0,50 \%$, $\hat{\sigma}_{sp} = 0,60 \%$, $\hat{\sigma}_{sys} = 0,30 \%$ und ein Bias von 2,00 % wurden in einer Simulation eingesetzt. Als Ergebnis zeigte sich mit einer Annahmewahrscheinlichkeit (α -Fehler) von 19 % an der Ablehngrenze eine starke Verzerrung, die 14 Prozentpunkte über dem eingesetzten α von 5 % lag.

4.2.6 Abhängigkeit der Power vom Bias

Die nachfolgende Abbildung 4.2-3 zeigt die Annahmewahrscheinlichkeit für einen Äquivalenztest dreier unterschiedlicher Situationen als stetige Funktion in Abhängigkeit des Bias, der von 0 bis über die Akzeptanzgrenze von 2 % hinaus verläuft.

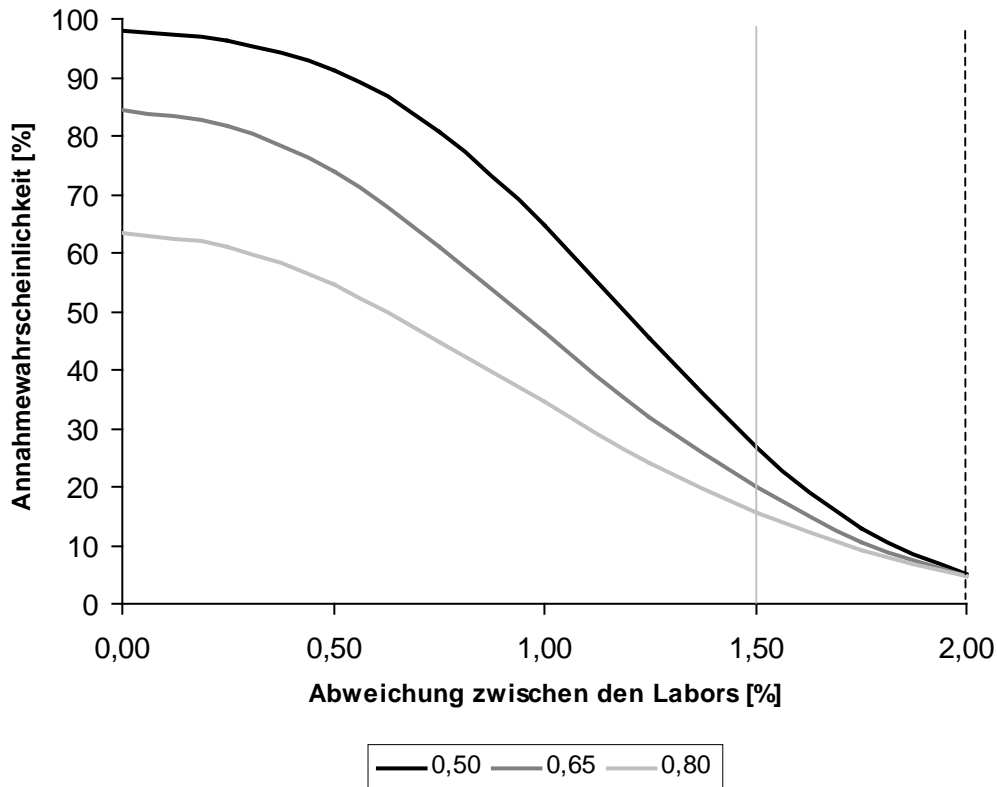


Abbildung 4.2-3: Annahmewahrscheinlichkeit des Äquivalenztestes bei verschiedenen Präzisionen.

Die Leistung des Äquivalenztestes ist in den drei abgebildeten Situationen präzisionsbedingt sehr unterschiedlich. Während bei einer RSD% von 0,80 % die Power von vornherein nur knapp über 60 % liegt, ist die Power bei 0,65 und 0,50 % anfänglich mit 85 bzw. 98 % ausreichend hoch.

In jedem Fall ist zu erkennen, dass selbst in den beiden Situationen mit 0,50 und 0,65 %, in denen bei einem kleinen Bias hohe Annahmewahrscheinlichkeiten vorliegen, diese Wahrscheinlichkeit stark abfällt (auf unter 30 %), sobald der Bias

1,5 % überschreitet. Mit anderen Worten: Ist die tatsächliche Abweichung zwischen den Labors mehr als halb so groß wie die Akzeptanzgrenze, dann ist die Power an dieser Stelle niedrig, unabhängig davon, ob die Power ohne Bias über 95 % betrug. Das gilt insbesondere für Situationen mit der Bedingung, dass die Power ohne Bias über 80 % liegen soll.

4.2.7 Leistungsfähigkeit des Äquivalenztestes unabhängig von der Akzeptanzgrenze (AkzG)

Bei der Betrachtung der Leistungsfähigkeit des Äquivalenztestes in Kapitel 4.2.3 sind alle Annahmewahrscheinlichkeiten für die verschiedenen Szenarien unter Verwendung einer AkzG von 2 % berechnet worden. Eine AkzG von 2 % wurde als Wert herangezogen, da sie eine übliche Größenordnung für Methodentransfers zur Gehaltsbestimmung von Arzneistoffen darstellt.

Sofern nicht vorgegeben, ist die Wahl der AkzG von vielen Parametern abhängig. Die laboreigene Präzision ist dafür ein wichtiger Faktor, um rationale AkzG festzulegen [60]. Bestimmte Methoden besitzen von vornherein eine geringe Systempräzision oder einen großen Beitrag durch Probenaufbereitung. Für solche Situationen ist eine AkzG von 2 % ungeeignet.

Außerdem kann es gewünscht sein, durch die Wahl aufgeweiteter AkzG den analytischen Aufwand durch eine kleinere Anzahl an Analytikern oder Serien bei vergleichbaren Annahmewahrscheinlichkeiten zu verringern.

Für alle möglichen Werte von Akzeptanzgrenzen einzelne Abbildungen zu erzeugen, wäre nicht möglich gewesen. Daher wurde als Variable, auf die sich die Leistungsfähigkeit des Äquivalenztestes bezieht, das prozentuale Verhältnis aus $\hat{\sigma}_{\bar{x}_s}$ und AkzG gewählt. Denn die Annahmewahrscheinlichkeit für einen bestimmten Bias und eine bestimmte Anzahl an Analytikern oder Serien zwischen verschiedenen Szenarien ist genau dann gleich, wenn der Quotient aus dem RSD%-Wert der $\hat{\sigma}_{\bar{x}_s}$ und der prozentualen AkzG gleich ist. Dafür wurde der Term **%AkzG** nach Gleichung 4.2-2 gebildet:

$$\%AkzG = \frac{\hat{\sigma}_{\bar{x}_{series}} \cdot 100}{AkzG} [\%] \quad (4.2-2)$$

Folgende Situation soll dies verdeutlichen:

Fall A: $\hat{\sigma}_{\bar{x}_s}$ der Labors = 0,60 %; AkzG = 2 %.

Die Berechnung der %AkzG nach Gl. 4.2-2 ergibt für diesen Fall 30 %.

Die Annahmewahrscheinlichkeit für Fall A beträgt 90 % für die Situation:

Bias zwischen den Labors = 0 und N = 3 Serien/Labor.

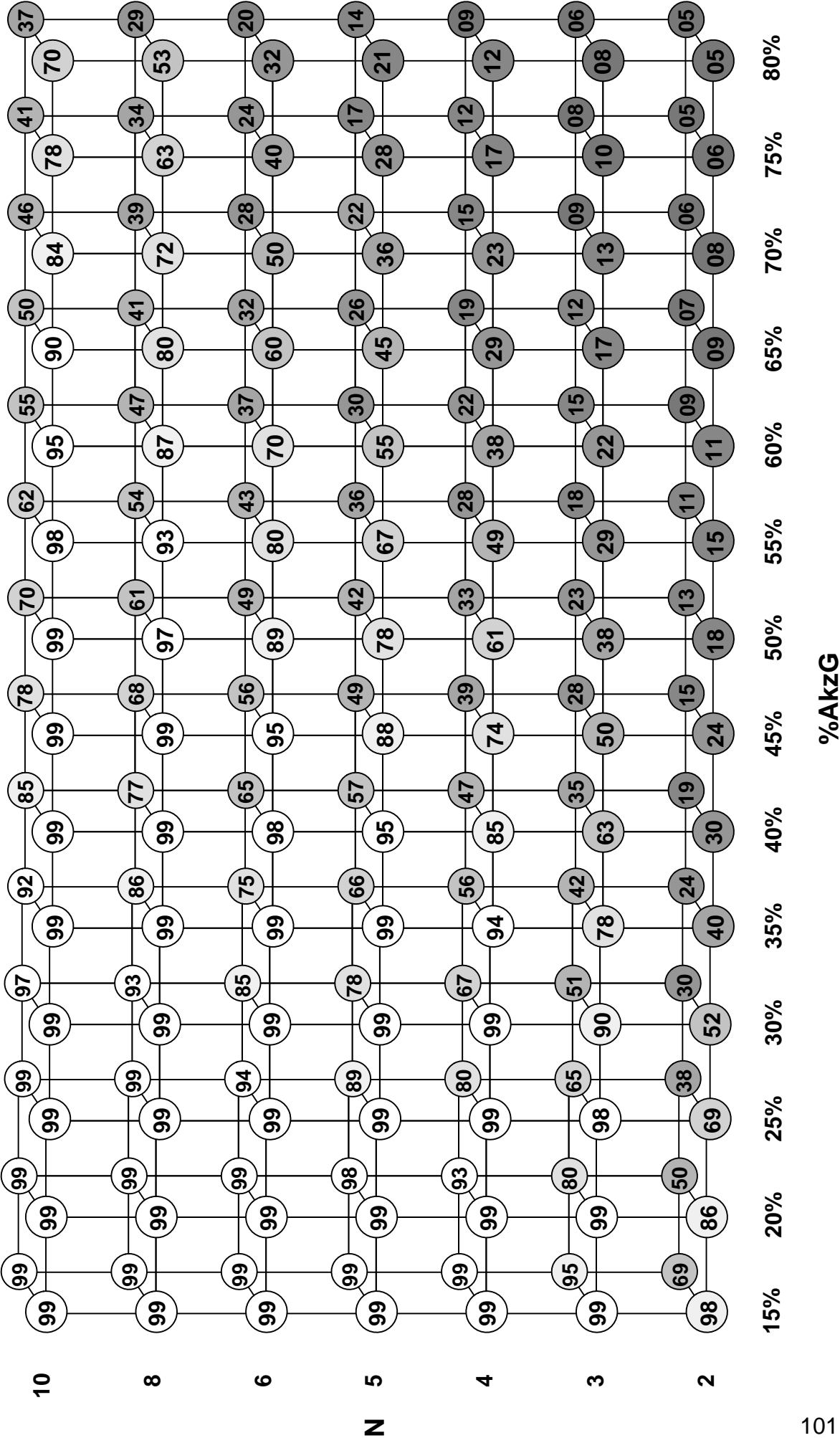
Fall B: $\hat{\sigma}_{\bar{x}_s}$ der Labors = 0,90 %; AkzG = 3 %.

Auch hier ergibt die Berechnung der %AkzG einen Wert von 30 %, weil die $\hat{\sigma}_{\bar{x}_s}$ ebenfalls betraglich 30 % des Wertes der AkzG ausmacht.

Die Annahmewahrscheinlichkeit beträgt auch hier 90 %, wenn man die gleiche Situation wie in Fall A (Bias = 0, N = 3) betrachtet.

Fall A und B besitzen also für alle Situationen (Bias/N) die gleiche Power, weil das Verhältnis von $\hat{\sigma}_{\bar{x}_s}$ und AkzG gleich ist.

Für %AkzG-Werte im Bereich von 15 bis 75 %, 2 bis 10 Serien und einem α von 0,05 sind die entsprechenden Simulationen mit dem Äquivalenztest durchgeführt worden. Die Ergebnisse sind als Annahmewahrscheinlichkeit in der nachfolgenden Abbildung 4.2-4 dargestellt.



(vorherige Seite) Abbildung 4.2-4: Die Werte in den Kreisen stehen für die Wahrscheinlichkeit, den Methodentransfer zurecht anzunehmen (Power). Der Bias, in der vorderen Ebene beträgt 0 % und in der hinteren 50 % der Akzeptanzgrenze. Auf der x-Achse ist der prozentuale Anteil der $\hat{\sigma}_{\bar{x}_s}$ an der Akzeptanzgrenze aufgetragen (berechenbar nach Gl. 4.2-2). Die y-Achse zeigt die Anzahl der an dem Transfer eingesetzten Serien je Labor. Die Irrtumswahrscheinlichkeit beträgt 5 %.

Bei Kenntnis bzw. Abschätzung der laboreigenen $\hat{\sigma}_{\bar{x}_s}$ und einer vorgegebenen oder wählbaren Akzeptanzgrenze kann die notwendige Anzahl einzusetzender Serien N für die gewünschte Power in Abbildung 4.2-4 abgelesen werden.

Ein einfacher Zusammenhang ist Abbildung 4.2-4 dahingehend zu entnehmen, dass für %AkzG-Werte in der Reihe 20, 30 und 40 % jeweils N = 2, 3 und 4 Serien notwendig sind, um eine Power von über 80 % zu sichern.

Für %AkzG-Werte von 50, 60 und 70 % wird der nichtlineare Zusammenhang zwischen der Anzahl von Serien und dem prozentualen Anteil deutlich, da hier bereits 6, 8 und 10 Serien notwendig sind.

4.3 Leistung beim Vergleich von Varianzen

4.3.1 Voraussetzung

Der angewandte F-Test setzt voraus, dass die Stichproben der Labors voneinander unabhängig sind. Der α -Fehler kann anderenfalls bei vorliegender Korrelation nicht auf die vorgegebenen 5 % kontrolliert werden. Wenn sich die geschätzten Varianzen beider Labors z.B. aus den Komponenten Charge und Methode zusammensetzen, dann können bei einer zu hohen Variation innerhalb der Charge die Ergebnisse korrelieren. Daher wird hier unterstellt, dass Faktoren, die sich auf beide Labors gleichartig auswirken, von vernachlässigbarer Größe sind (z.B. unter 5 % der gesamten Varianz).

4.3.2 Auswahl der Testung

Die Testung zweier Varianzen auf Gleichheit für den Methodentransfer basiert auf dem Vorschlag des USP-Kapitels <1010> [46, 61]. Darin wird ein 90%-Konfidenzintervall um das im Rahmen eines F-Testes geschätzte Verhältnis der Varianzen der beiden beteiligten Labors gebildet. Anschließend wird die obere Grenze des Intervalls mit einer vorher festgelegten Akzeptanzgrenze verglichen (s. 2.6.9).

4.3.3 Leistungsfähigkeit des Äquivalenztestes

Das Risiko, fälschlicherweise ein höheres Verhältnis als vorgegeben zu akzeptieren, wird durch den Äquivalenztest auf 5 % festgelegt (s. 2.6.10). Die Wahrscheinlichkeit der korrekten Annahme des Transfers (Power) sowie das Risiko der ungerechtfertigten Ablehnung eines erfolgreichen Transfers sind vom wahren, unbekannten Verhältnis der beteiligten Varianzen der Labors, von der Höhe der Akzeptanzgrenze und von der Stichprobengröße abhängig. Die Beziehungen zwischen diesen Größen sind direkt berechenbar [46]. Für typische Varianzbeträge und Akzeptanzgrenzen sind nach der in Abschnitt 7.2.2 gegebenen Formel die erhaltenen Annahmewahrscheinlichkeiten berechnet worden.

4.3.4 Ergebnisse für die USP <1010> Methode

Eine Übersicht über die berechneten Annahmewahrscheinlichkeiten für verschiedene Akzeptanzgrenzen und Stichprobengrößen wird in Abbildung 4.3-1 gegeben.

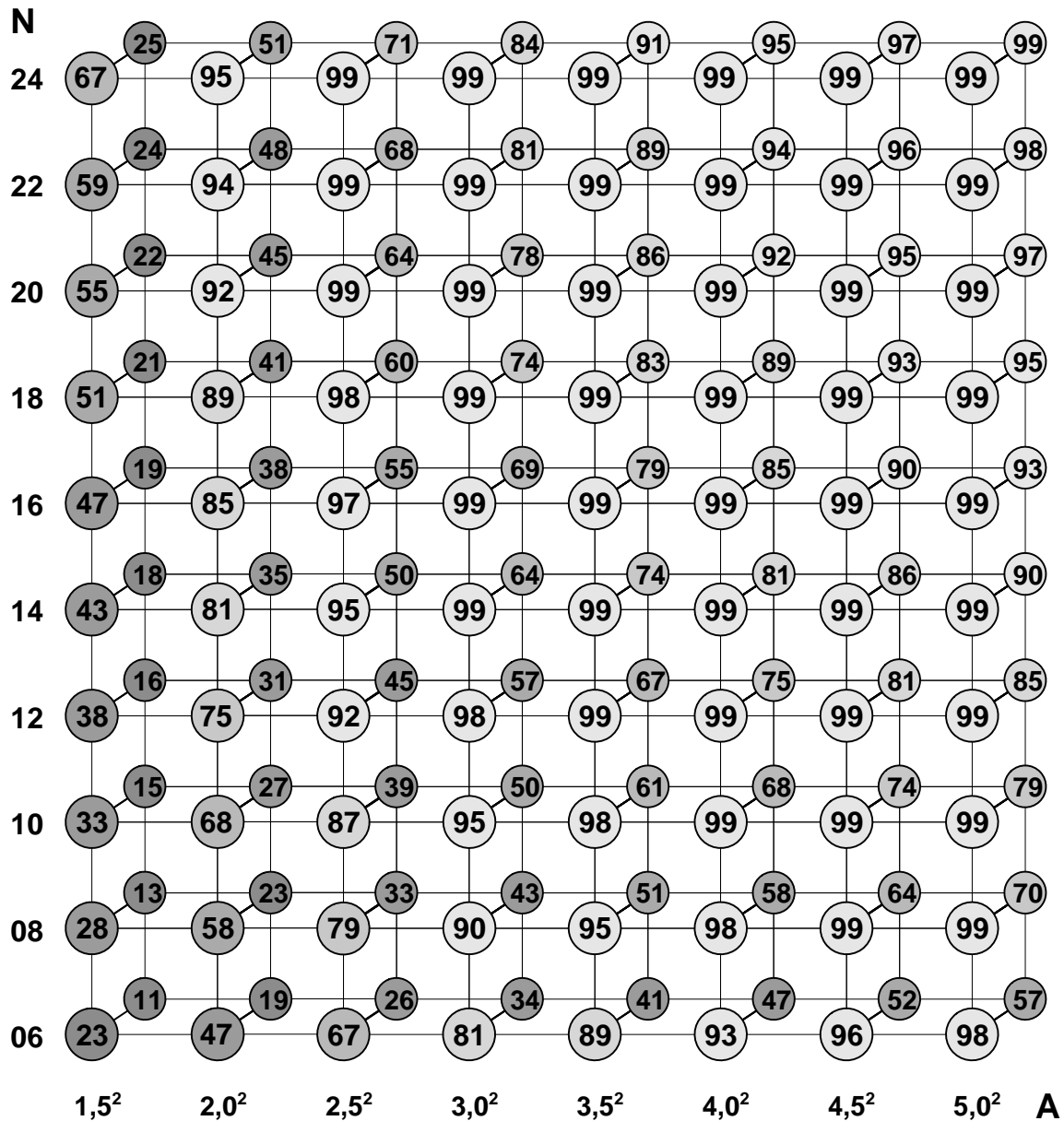


Abbildung 4.3-1: Power verschiedener Transfersituationen. Auf der x-Achse ist die vorher festgelegte obere Akzeptanzgrenze A gegeben. Die y-Achse enthält die Stichprobengröße N beider Labors. Die Kreise in der vorderen Ebene geben die entsprechende Power an, wenn keine Abweichung zwischen den Labors besteht (Verhältnis der Varianzen = 1). Das tatsächliche Verhältnis der Varianzen für die hintere Ebene ist halb so hoch wie die jeweilige obere Akzeptanzgrenze der Spalte. Diese berechnet sich nach Gl. 4.3-1 aus Abschnitt 7.2.3.

Die exemplarische obere Grenze A von $2,0^2$ ($4,0$) benötigt eine Stichprobengröße von mindestens 14 um eine Power von mehr als 80 % zu erreichen. Die Wahl der Akzeptanzgrenze beeinflusst stark die Power. Eine kleine obere Grenze von $A = 2,25$ ($1,5^2$) ist nicht praktikabel, da selbst bei einer Stichprobengröße von $n = 24$ (die als hoch angesehen werden kann) nur eine Chance von 62 % für die Annahme des Transfers besteht, wenn die verglichenen Varianzen in Wahrheit gleich sind. Ein Verhältnis zweier Varianzen zueinander von 2,25 liegt dann vor, wenn die Standardabweichung des Empfängerlabors in Wahrheit das 1,5-fache der des Ursprungslabors beträgt.

Des Weiteren ist zu sehen, dass eine obere Grenze von 16 (entsprechend dem Vorliegen einer vierfachen Standardabweichung des Empfängerlabors) sehr leicht zu kontrollieren ist, denn bereits eine Stichprobengröße von sechs reicht aus, um mit hoher Wahrscheinlichkeit den Transfer zurecht anzunehmen. Derartig hohe Akzeptanzgrenzen sind normalerweise nur dann sinnvoll, wenn das Ursprungslabor eine sehr hohe Präzision aufweist. Bei einer RSD% von 0,20 % wäre ein bis zu vierfacher RSD%-Wert für das Empfängerlabor noch ohne weiteres akzeptabel. Üblicherweise legen erfahrene Analytiker ihr Augenmerk nicht auf sehr kleine Standardabweichungen. Ein Wert von 0,20 % wird schwierig zu reproduzieren sein, selbst wenn mit einer hohen Datenzahl gemessen wird. An weiteren Tagen könnte eine kleine zusätzliche Variation, die vorher nicht in Erscheinung getreten ist, die Variation leicht verdoppeln.

Betrachtet man die obere Grenze A als Quotienten der Quadrate zweier Standardabweichungen, so lässt sich $A = 4/1$ als $2^2/1^2$ schreiben. Dann führt eine Erhöhung der oberen Grenze um den Wert 0,25 (also: $x + 0,25$)² im Intervall zwischen $A = 2^2$ und 3^2 dazu, dass bei einer Verringerung der Stichprobengröße N um 2 trotzdem eine Power von über 80 % beibehalten wird (s. gestrichelte Linie in Abb. 4.3-1).

Die oben aufgeführten Annahmewahrscheinlichkeiten sind nur dann gültig, wenn die wahren Varianzen des Ursprungs- und des Empfängerlabors gleich sind. Es ist ebenso wichtig zu schauen, welche Power erwartet werden kann, wenn die Varianzen unterschiedlich sind aber noch innerhalb des Akzeptanzbereichs liegen. Ist das tatsächliche Verhältnis der Varianzen R gleich der Akzeptanzgrenze A , dann beträgt die Annahmewahrscheinlichkeit exakt 5 %, gemäß der Konstruktion des Konfidenzintervalls.

Um ablesen zu können, welche Power zwischen diesen beiden Extremen erwartet werden kann, ist die Power für die Hälfte des noch zulässigen Verhältnisses berechnet worden und in der hinteren Ebene der Abbildung 4.3-1 dargestellt. Dieses Verhältnis liegt zwischen 1 und der Akzeptanzgrenze A. Die Stelle, an der die Hälfte des zulässigen Verhältnisses vorliegt, berechnet sich gemäß Gleichung 4.3-1:

$$R = \left[1 + \frac{(\sqrt{A} - 1)}{2} \right]^2 \quad (4.3-1)$$

Ein Rechenbeispiel ist in Abschnitt 7.2.3 gegeben. Eine ausreichende Power von 80 % oder höher kann nur mit einer ausreichend großen Stichprobenzahl erreicht werden. Diese Datenzahlen sind deutlich höher verglichen mit denen für den Fall gleicher Varianzen. Wie an Abbildung 4.3-1 zu erkennen ist, bedarf eine Akzeptanzgrenze $A = 3^2$ ein N von 22. Um N zu verringern, müsste man eine tolerantere Grenze wählen (z.B. $A = 4^2$ und $N = 14$).

Werte für A wie 16 bzw. 4^2 dürften oft die bessere Wahl sein. Man betrachte $A = 16$, $N = 6$, so fällt die Power stark ab, wenn ein wirklich großer Unterschied zwischen den Varianzen vorliegt. Das bedeutet auch, dass der Test in den meisten Fällen den Transfer ablehnt, wenn das wahre Verhältnis zwischen 7 und 16 liegt. Denn eine Grenze von 16 bedeutet nicht, dass es wahrscheinlich ist, für wahre Quotienten nahe der Grenze, den Transfer anzunehmen.

4.4 Testung der Standardabweichung gegen eine Akzeptanzgrenze

Die Testung des oberen Konfidenzintervalls der geschätzten Standardabweichung gegen eine Akzeptanzgrenze kontrolliert ebenfalls das Risiko, fälschlicherweise eine höhere Standardabweichung als vorgegeben zu akzeptieren, auf 5 % (s. 2.6-11). Die Wahrscheinlichkeit der korrekten Annahme des Transfers (Power) sowie das Risiko der ungerechtfertigten Ablehnung eines erfolgreichen Transfers hängen von der wahren, unbekannten Standardabweichung, der Höhe der Akzeptanzgrenze und von der Stichprobengröße ab. In Abbildung 4.4-1 sind für gängige Stichprobengrößen gegenüber eines Anteils der Präzision bezogen auf die Akzeptanzgrenze zwischen 30 und 100 % Annahmewahrscheinlichkeiten berechnet und dargestellt worden.

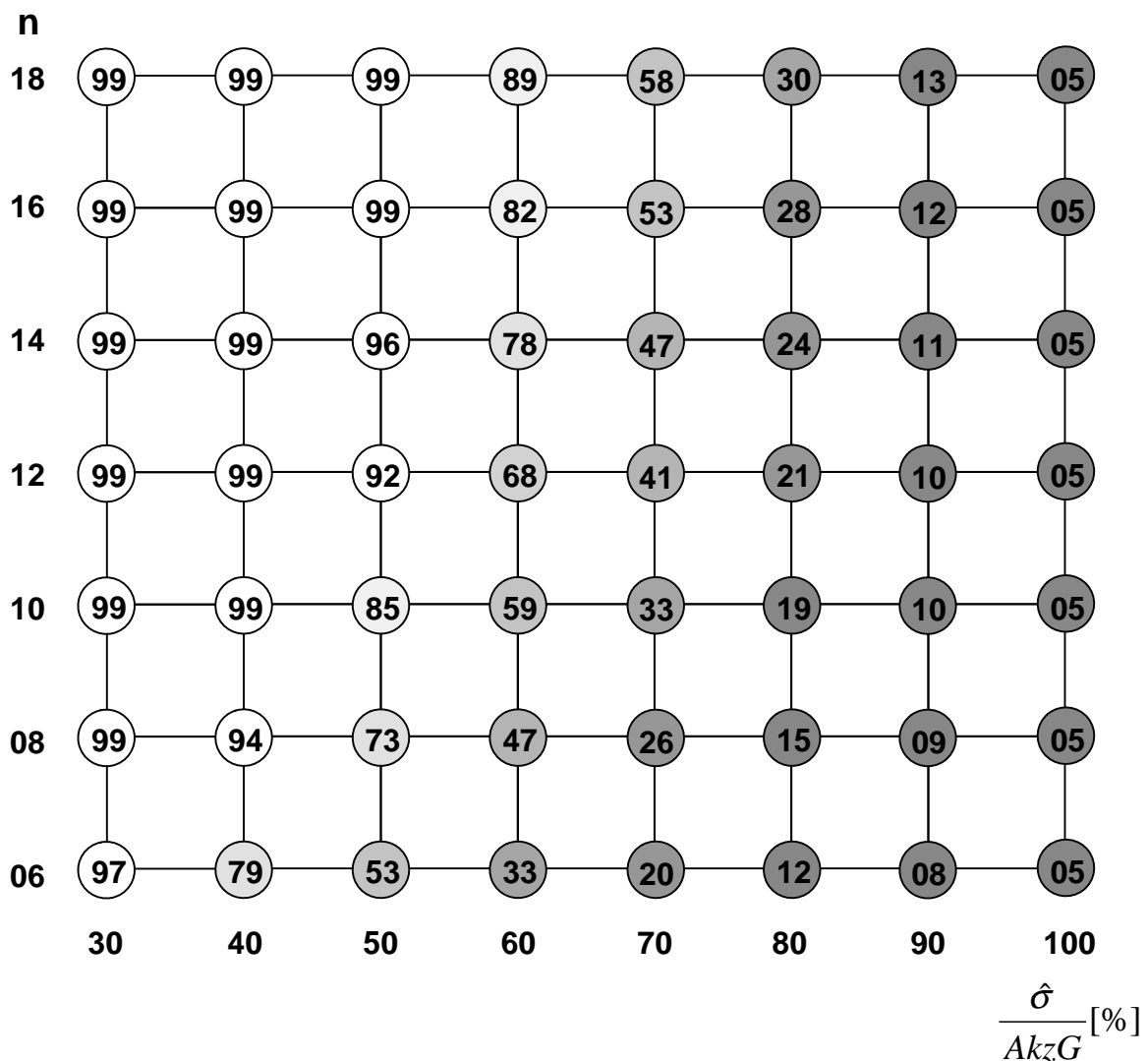


Abbildung 4.4-1 (vorherige Seite): Annahmewahrscheinlichkeit in Abhängigkeit der Stichprobengröße und des prozentualen Anteils der tatsächlichen Präzision bezogen auf die Akzeptanzgrenze. Diese Übersicht erlaubt die Abschätzung der notwendigen Anzahl an Freiheitsgraden für unterschiedliche Situationen.

Ist der Anteil der tatsächlichen (unbekannten) Präzision bezogen auf die Akzeptanzgrenze kleiner als 30 %, dann reicht eine Stichprobengröße von $n = 6$ aus, um mit hoher Wahrscheinlichkeit eine solche Testung zurecht anzunehmen. Oberhalb eines relativen Anteils von 60 % steigt der Aufwand der Testung stark an ($n > 20$). Ist die Präzision, von der ausgegangen wird, halb so hoch wie die Akzeptanzgrenze, dann sichert eine Stichprobengröße von 10 eine ausreichend hohe Annahmewahrscheinlichkeit (85 %).

5 Zusammenfassung

Die Untersuchungen zur Präzision der HPLC-Gehaltsbestimmung von Glibenclamid ergeben eine Reihe neuer Erkenntnisse über die Entwicklung und die Zusammensetzung der Wiederhol-Standardabweichung in den verschiedenen Konzentrationsbereichen. Abbildung 5.1 stellt die gewonnenen Erkenntnisse schematisch dar.

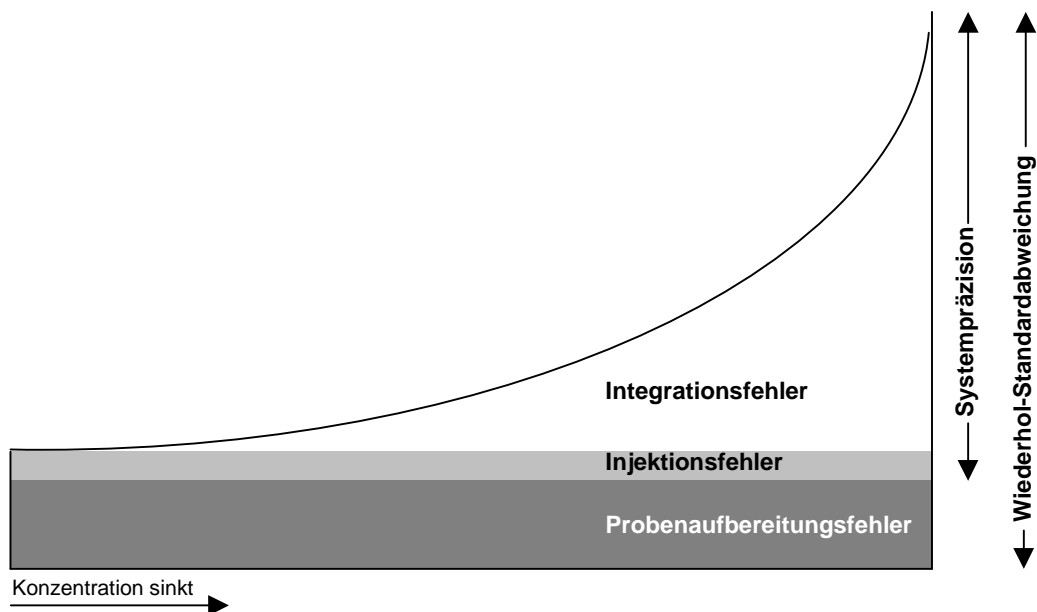


Abbildung 5-1: Vereinfachte Darstellung der Entwicklung und Zusammensetzung der Wiederhol-Standardabweichung im Verlauf gegen kleine Konzentrationen.

Die Wiederhol-Standardabweichung verläuft in Reihe A der durchgeführten Messungen beginnend vom Arbeitsbereich der Gehaltsbestimmung (120 %) bis zu einer Konzentration von 10 % konstant. Aus den einzelnen Bestimmungen der jeweiligen Konzentrationen lässt sich die Wiederhol-Standardabweichung für dieses Intervall zu einem Wert von 0,70 % zusammenfassen. Analog dazu ergibt sich für die ebenfalls in diesem Bereich konstante Systempräzision ein zusammengefasster Wert von 0,31 %. Da sich die Wiederhol-Standardabweichung aus den unabhängigen Variationskomponenten Systempräzision und Probenaufbereitung zusammensetzt, verbleibt für die Probenaufbereitung ein Betrag von 0,63 %. Die Probenaufbereitung stellt damit den Hauptteil an der Wiederhol-Standardabweichung dar (ca. 80 %). Der Wert der

Systempräzision von 0,31 % kann ganz dem Injektionsfehler zugeordnet werden, da der Integrationsfehler in diesem Bereich von untergeordneter Bedeutung ist.

Im Konzentrationsbereich von 10 bis 0,25 % steigt der Wert für die Wiederhol-Standardabweichung der Methode linear gegen den Logarithmus der Konzentration an. Da der Beitrag zur Wiederhol-Standardabweichung durch die Probenaufbereitung konstant ist, nimmt die Systempräzision einen immer größeren Anteil an der Wiederhol-Standardabweichung ein. Die Zunahme der Systempräzision ist durch den Integrationsfehler als einzige variable Fehlerquelle bedingt. Der Integrationsfehler übersteigt am Ende dieses Bereichs betragsmäßig die Variation durch Probenaufbereitung und wird damit zur Hauptfehlerquelle.

Im Konzentrationsbereich unterhalb von 0,25 % besteht die Wiederhol-Standardabweichung fast ausschließlich aus der Systempräzision und damit dem Integrationsfehler, der als dominierende Fehlerquelle letztlich den exponentiellen Anstieg bewirkt.

Die Verbesserung der Präzision durch die Erhöhung der Empfindlichkeit der Methode in Reihe B belegt die Dominanz des Integrationsfehlers in den untersten Konzentrationsbereichen.

Je näher sich die Systempräzision betragsmäßig der Wiederhol-Standardabweichung annähert, desto bedeutender wird der Integrationsfehler. Zwei sinnvolle Möglichkeiten zur Senkung relativ hoher RSD%-Werte ($> 2\%$) sind in solchen Fällen entweder die Erhöhung der Empfindlichkeit der Methode oder Mehrfachinjektionen aus den Proben.

Die Arbeit zeigt, dass in der Regel entweder der Probenaufbereitungsfehler oder der Integrationsfehler die Hauptfehlerquelle darstellt. Der Injektionsfehler hingegen ist immer von untergeordneter Bedeutung. Durch den Vergleich von Wiederhol-Standardabweichung und Systempräzision lässt sich die jeweilige Hauptfehlerquelle ermitteln.

In Reihe C sind die Nebenprodukte (**3-6**) im Konzentrationsbereich von 1,0 bis 0,025 % untersucht worden. Bezüglich des Verlaufs der Präzision weisen Sie ein zu Glibenclamid vergleichbares Verhalten auf, wenn man die unterschiedlichen Parameter wie Retentionszeit oder spezifische Absorption der Stoffe berücksichtigt.

Die Berechnung der Mehrtages-Standardabweichung mit den Werten des Bereichs von 100 bis 10 % ergibt einen RSD%-Wert von 0,88 %. Verglichen mit

der Variation innerhalb der Serien (65 %) liefert die Variation zwischen den Serien mit einem Anteil von 35 % an der Mehrtages-Standardabweichung einen geringeren Beitrag, was dem relativ engen betrachteten Zeitraum zuzuschreiben ist.

Die Wiederfindung der Gehaltsbestimmung aus Reihe A bleibt bis zu einer Konzentration von 0,075 % konstant und systematisch erhöht. Für die Nebenprodukte in Reihe C ergeben sich die Abweichungen von 100 % aus dem Problem des Setzens der Basislinien.

Die Übereinstimmung von statistischem und empirischem Vertrauensbereich anhand der Mindestanforderungen der ICH-Richtlinie Q2B belegt auch die korrekte Normalverteilungsannahme der empirischen HPLC-Daten des betrachteten Bereichs.

Die Leistungsfähigkeit des Äquivalenztestes für den Vergleich von Mittelwerten bzw. von Varianzen im Rahmen analytischer Methodentransfers konnte eingehend beurteilt werden.

Für den Vergleich von Mittelwerten wurde für das ISPE-Versuchsdesign anhand von Simulationen ein Diagramm (s. Abb. 4.2-2) erstellt, das die Ermittlung der Power in Abhängigkeit von Stichprobengröße und Präzision ermöglicht. Dabei liefert das feste ISPE-Versuchsdesign einen ausreichenden Wert für die Power ($> 80\%$), wenn die Gesamtstandardabweichung der beteiligten Analytiker entweder unter 0,4 % liegt oder mehr als zwei Analytiker je Labor eingesetzt werden - und zwar je Zunahme der Gesamtstandardabweichung um 0,2%-Punkte mindestens ein weiterer Analytiker. Bei zwei Analytikern und einer Variation von z.B. 0,62 % beträgt die Power lediglich 50 %.

Mit Hilfe zusätzlicher Simulationen wurde ein Diagramm (s. Abb. 4.2-4) zum Ablesen der zu erwartenden Power erstellt, das von der absoluten Höhe der Akzeptanzgrenze und der Präzision unabhängig ist. Ist die Präzision im Vergleich zur Akzeptanzgrenze betragsmäßig gering (kleiner %AkzG-Wert), lässt sich eine einfache Regel ableiten: Für %AkzG-Werte von 20, 30 oder 40 % ist eine Anzahl von 2, 3 oder 4 Serien notwendig, um eine Power von über 80 % zu erreichen. Sind die eigenen Variationskomponenten bekannt oder hinreichend exakt abschätzbar, so kann durch diese Untersuchung bei bekannter Akzeptanzgrenze leicht die notwendige Anzahl an Serien ermittelt werden.

Bezüglich der Leistungsfähigkeit des Äquivalenztestes für den Vergleich von Varianzen wurde ein Diagramm (s. Abb. 4.3-1) erstellt, an dem die Power für verschiedene Stichprobenumfänge und verschiedene Akzeptanzgrenzen abgelesen werden kann. Der wahrscheinlichste Fall der Wahl einer Akzeptanzgrenze A liegt im Bereich von 4 bis 9 bzw. $2^2/1^2$ bis $3^2/1^2$ (als Quotient der Quadrate zweier Standardabweichungen). Innerhalb dieses Intervalls gilt für den Fall tatsächlich gleicher Varianzen: Für jede weitere Erhöhung der oberen Akzeptanzgrenze A um 0,25 (z.B. $2^2 \Rightarrow 2,25^2$) kann man die Stichprobenzahl jeweils um 2 verringern und behält weiterhin eine Power von über 80 % bei.

Liegt ein akzeptabler Unterschied zwischen den Varianzen vor, sind die Werte für die Power deutlich geringer. Ein Wert von 80 % ist dann nur noch mit einer sehr hohen Stichprobenzahl erreichbar.

Es wird ersichtlich, dass der sinnvolle Einsatz des Äquivalenztestes für Varianzen einen im Vergleich zum Einsatz bei Mittelwerten höheren Aufwand nach sich zieht. Die Betrachtungen zeigen, dass ein festes Versuchsdesign mit einer festen Stichprobengröße nicht grundsätzlich zum Erfolg führt - insbesondere bei Unkenntnis über die Beträge der vorhandenen Varianzkomponenten. Um auch in Grenzfällen eine ausreichende Power zu erhalten, sollte eher eine höhere als eine knapp kalkulierte Stichprobenzahl gewählt werden, da die Power mit der Stichprobengröße steigt.

6 HPLC-experimenteller Teil

6.1 Geräte

Die Zusammensetzung der eingesetzten HPLC-Anlage aus den einzelnen Komponenten kann aus Tabelle 6.1-1 entnommen werden.

Komponente	Typ	Hersteller
Fließmittelpumpe	L 6200 A	Merck
Autosampler	AS 2000 A	Merck
Säulenofen	T1	Techlab
Säule	Superspher 100 RP18, endcapped, 4µm Partikelgröße, 125x4mm	Merck
Diodenarraydetektor	L 7450	Merck
Interface	D 6000 A	Merck
Software	D 7000 HSM	Merck

Tabelle 6.1-1: Zusammensetzung der HPLC-Anlage

Für die Probenaufbereitung kamen folgende Geräte zum Einsatz: Substanzeinwaagen im Bereich von 0,375 bis 12,5 mg wurden an einer Sartorius Mikro Waage (Modell 4431) vorgenommen. Größere Einwaagen erfolgten an einer Sartorius Waage (Modell BA 210 S). Alle 1 mL-Verdünnungsschritte wurden mit einer fixen 1 mL Eppendorf Pipette durchgeführt. Als Zentrifuge diente das Modell 5417C (Eppendorf) und als Ultraschallbad das Modell Sonorex RK100 (Bandelin).

6.2 Analyten

Tabelle 6.2-1 zeigt die eingesetzten Analyten (**1-6**). Die Substanzen wurden von Aventis (Frankfurt, Deutschland) gestellt.

Nr.	Substanz	Reinheit [%]
(1)	1-[[4-[2-[(5-Chlor-2-methoxybenzoyl)amino]ethyl]phenyl]-sulfonyl]-3-cyclohexylharnstoff	>99,9
(2)	5-Chlor-2-methoxy-N-[2-(4-sulfamoylphenyl)ethyl]-benzamid	>99,0
(3)	Methyl[[4-[2-[(5-chlor-2-methoxybenzoyl)amino]ethyl]-phenyl]sulfonyl]-carbamate	>95,6
(4)	1-[[4-[2-[(3-Ethyl-4-methyl-2-oxo-2,5-dihydro-1H-pyrrol-1-yl)-carbonyl]amino]ethyl]phenyl]-sulfonyl]-3-(trans-4-methylcyclohexyl)harnstoff	>99,7
(5)	Ethyl[[4-[2-[(3-ethyl-4-methyl-2-oxo-2,5-dihydro-1H-pyrrol-1-yl)-carbonyl]amino]ethyl]phenyl]sulfonyl]-carbamate	>99,5
(6)	1-[[3-[2-[(3-Ethyl-4-methyl-2-oxo-2,5-dihydro-1H-pyrrol-1-yl)-carbonyl]amino]ethyl]phenyl]-sulfonyl]-3-(trans-4-methylcyclohexyl)harnstoff	>98,8

Tabelle 6.2-1: Die Analyten (**1-6**). (**1**) = Glibenclamid und (**4**) = Glimepirid.

Die entsprechenden Strukturformeln sind in nachfolgender Abbildung 6.2-1 zu sehen, wobei (1-3) auf der Struktur von Glibenclamid und (4-6) auf der des Glimepirids basieren.

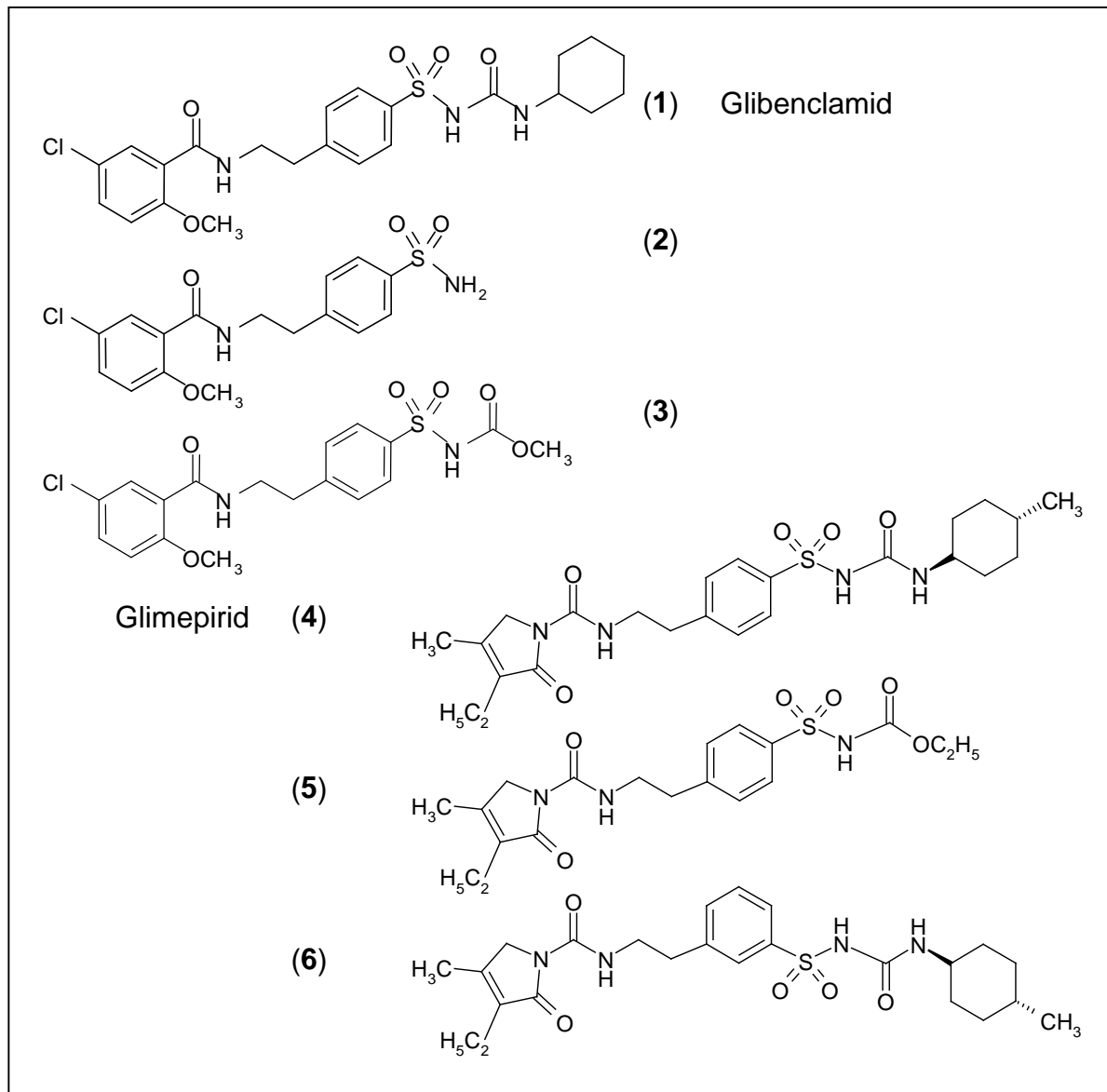


Abbildung 6.2-1: Glibenclamid (1) neben den Modellverunreinigungen (2-6).

6.3 Reagenzien

Als Reagenzien für die mobile Phase, das Lösungsmittel und die Tablettenmatrix sind Acetonitril für HPLC (Acros organics, Gelnhausen, Deutschland und von Fischer Scientific, Schwerte, Deutschland), Wasser (Lichrosolv®), KH_2PO_4 , $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$, $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, Phosphorsäure 85 %, Mikrokristalline Zellulose, Polyvidone 25000, Laktosemonohydrat, Natriumstärkeglykolat, Magnesiumstearat (alle Merck, Darmstadt, Deutschland), sowie Gelbes Eisenoxid (Caelo, Hilden, Deutschland) eingesetzt worden.

6.4 Methoden

6.4.1 Reihe A

Die HPLC-Methode für die Reihe A bestand aus folgenden Einstellungen:

Flussrate = 1,25 mL/min

Mobile Phase = Acetonitril-Puffer (45:55, v/v)

Temperatur des Säulenofens = 35 °C

Injektionsvolumen = 10 µL

Detektionswellenlänge für (1) = 210 nm

Länge eines Laufes = 10 min

Der Konzentrationsbereich unterhalb von 1,0 % der Reihe A wurde neben dem sogenannten „normal mode“ des Diodenarraydetektors auch mit dem „low mode“ vermessen. Der „low mode“ erzielt eine 10-fach höhere digitale Auflösung, was aber keinen Effekt auf die Systempräzision nach sich zog.

6.4.2 Reihe B

Die HPLC-Methode für die Reihe B unterschied sich von der für die Reihe A nur in dem erhöhten Injektionsvolumen von 100 µL.

6.4.3 Reihe C

In Reihe C wurde die Methode der Reihe A eingesetzt. Die Detektionswellenlängen für die zusätzlichen Analyten waren für (3) und (6) 210 nm sowie für (4) und (5) 228 nm. Abweichend von Reihe A betrug die Länge eines Laufes 12,5 min statt 10 min, da mit (4) und (6) zwei Substanzen mit im Vergleich zu (1) längeren Retentionszeiten vermessen wurden.

6.5 Messserien

Jede Messserie setzte sich aus einer Injektionssequenz und den zugehörigen Spülschritten auf folgende Weise zusammen [62, 63]:

1. Spülschritt: Mobile Phase	(0,50 mL/min - 40 min)
2. Spülschritt: Mobile Phase	(1,25 mL/min - 20 min)
3. Injektionssequenz (siehe 6.6.)	
4. Spülschritt: Acetonitril-Puffer, 5:95, v/v	(0,50 mL/min - 30 min)
5. Spülschritt: Acetonitril	(0,50 mL/min - 30 min)

6.6 Injektionssequenzen

Die Injektionssequenz für die einzelnen Konzentrationsniveaus der Reihen A und C bestand aus mindestens sechs Einzelinjektionen der Standards und einer anschließenden Leerinjektion, um Substanzverschleppung (carry over) zu vermeiden [64]. Es folgten sechs Einzelinjektionen der unabhängig aufbereiteten Proben, eine weitere Leerinjektion und schließlich die 6-fache Wiederholinjektion einer Probe. Im Unterschied dazu wurden in der Injektionssequenz der Reihe B die Proben anhand von 6 Doppelinjektionen vermessen.

6.7 Probenaufbereitung

6.7.1 Mobile Phase

Zur Herstellung der mobilen Phase (Acetonitril-Puffer, 45:55, v/v) wurden 650 mg $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$ in 550 g Wasser gelöst und anschließend zwei Tropfen Phosphorsäure 85 % sowie 351,5 g Acetonitril hinzugefügt, so dass sich bei 20 °C 1000 mL ergeben. Danach erfolgte die Entgasung der mobilen Phase im Ultraschallbad [65]. Die resultierenden pH-Werte nach der Zugabe der zwei Tropfen Phosphorsäure 85 % lagen zwischen 3,0 und 3,3. Schwankungen des pH-Wertes dieser Größenordnung zeigten keinen Einfluss auf die Trennung.

6.7.2 Lösungsmittel

Für das Lösungsmittel (Acetonitril-Puffer, 80:20, v/v) wurden 190,0 g Wasser, 10,0 g Phosphatpuffer pH 7 und 626,0 g Acetonitril eingewogen (entsprechend 1000 mL). Der letzte Verdünnungsschritt in der Reihe B (siehe 3.1.4) ist mit Lösungsmittel der Zusammensetzung (Acetonitril-Puffer, 40:60, v/v) durchgeführt worden. Dabei setzen sich 1000 mL aus 595,0 g Wasser; 5,0 g Phosphatpuffer pH 7 und 312,4 g Acetonitril zusammen. Der Phosphatpuffer pH 7 wurde hergestellt, indem 0,88 g KH_2PO_4 und 1,82 g $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$ in 250 mL Wasser gelöst wurden.

6.7.3 Tablettenmatrix

Die Tablettenmatrix diente zur Simulation der Gehaltsbestimmung der Arzneiform *Tablette* in den Reihen A und B. Dazu wurden alle Proben dieser Reihen vor dem letzten Verdünnungsschritt mit je 90 mg dieser Matrix versetzt. Die Zusammensetzung von 90 mg der Matrix geht aus Tabelle 6.7-1 hervor. Nach den Einwaagen wurde die Tablettenmatrix mit einem Mörser homogenisiert.

Substanz	Einwaage
Laktosemonohydrat	74,6
Natriumstärkeglykolat	4,0
Polyvidone 25000	0,5
Mikrokristalline Zellulose	10,0
Magnesiumstearat	0,5
Gelbes Eisenoxid	0,4

Tabelle 6.7-1: Zusammensetzung der Matrix.

6.7.4 Einwaage der Proben

Die jeweilige Substanz wurde zuerst in einen vortarierten Zinnbecher geeigneter Größe eingewogen. Danach ist die Substanz über einen Trichter in den vorgesehenen Messkolben gekippt und anschließend der Zinnbecher zurückgewogen worden. Die Einwaage ergab sich daher aus der Differenz des befüllten Zinnbechers zu dem leeren Zinnbecher.

6.7.5 Standards

Die jeweiligen Standards und die Proben sind immer am selben Tag hergestellt worden. Für den 100%-Standard wurden 5 mg Glibenclamid zu 25 mL mit Lösungsmittel in einem Messkolben aufgefüllt, was einer Konzentration von 0,2 mg/mL entspricht. Für einen 1%-Standard mit einer Konzentration von 0,002 mg/mL wurden entsprechend 12,5 mg zu 250 mL aufgefüllt und anschließend 1 mL dieser Lösung zu 25 mL verdünnt. Ein 0,1%-Standard zog noch einen weiteren 1:25-Verdünnungsschritt nach sich.

6.7.6 Proben der Reihe A

Für die Reihe A sind Glibenclamid-Proben in folgenden Konzentrationen hergestellt worden: 120; 100; 80; 60; 40; 20; 10; 7,5; 5,0; 2,5; 1,0; 0,75; 0,50; 0,25; 0,10; 0,075; 0,050 und 0,025 %. Die einzelnen Probenaufbereitungsschritte zur Herstellung dieser Konzentrationen können aus der nachfolgenden Tabelle 6.7-2

entnommen werden. Die 90 mg der Tablettenmatrix wurden stets vor dem letzten Verdünnungsschritt im Zielmesskolben vorgelegt. Anschließend erfolgte die Behandlung der Suspensionen für 15 min im Ultraschallbad. Danach ist die Probe für drei Minuten bei 14000 U/min zentrifugiert und letztlich der Überstand in entsprechende Vials abgefüllt worden.

Reihe	A		
Konzentration [%]	120 – 7,5	7,5 – 0,25	0,1 – 0,025
Einwaage [mg]	6,0 – 0,375	93,75 – 3,125	12,5 – 3,125
Stammlösung [mL]	25	250	100
1. Verdünnung	keine	1 : 25	1 : 25
2. Verdünnung	keine	keine	1 : 25

Tabelle 6.7-2: Probenaufbereitungsschritte der Reihe A.

6.7.7 Proben der Reihe B

Die Reihe B beinhaltet die drei Konzentrationslevel 0,1; 0,05 und 0,01 %. Die Probenaufbereitungsschritte sind der Tabelle 6.7-3 zu entnehmen. Im Unterschied zu den anderen Reihen wurde in Reihe B der 2. Verdünnungsschritt mit dem Lösungsmittel (Acetonitril-Puffer, 40:60, v/v) durchgeführt.

Reihe	B
Konzentration [%]	0,1 – 0,01
Einwaage [mg]	12,5 – 1,25
Stammlösung [mL]	100
1. Verdünnung	1 : 25
2. Verdünnung	1 : 25

Tabelle 6.7-3: Probenaufbereitungsschritte der Reihe B.

6.7.8 Proben der Reihe C

Jede Probenlösung der Reihe C bestand aus einem Gemisch der vier Modellverunreinigungen (**3-6**) in Konzentrationen zwischen 1,0 und 0,025 % in Anwesenheit von (**1**) als Matrix. Hierdurch sollte die quantitative Analyse von Verunreinigungen in einem Arzneistoff nachgestellt werden. Die Substanzen (**3-6**)

lagen jeweils in verschiedenen Konzentrationen vor, die unabhängig voneinander hergestellt wurden. Die Konzentrationsstufen waren: 1,0; 0,75; 0,5; 0,25; 0,10; 0,075; 0,050 und 0,025 %. Um eine Mindesteinwaage von mehr als 1 mg zu gewährleisten, wurden die Probenaufbereitungsschritte im Vergleich zur Reihe A etwas modifiziert. (1) ist vor jedem letzten Verdünnungsschritt in einer Konzentration von 100 % zugesetzt worden. Die jeweiligen Verdünnungsschritte sind in Tabelle 6.7-4 und die relative Zusammensetzung der Probenlösungen in Tabelle 6.7-5 aufgeführt. Ein exemplarischer Probenaufbereitungsplan ist im Anhang 10.1 dargestellt.

Reihe	C		
Konzentration [%]	1,0 – 0,25	0,1 – 0,05	0,025
Einwaage [mg]	5,0 – 1,25	3,125 – 1,563	3,125
Stammlösung [mL]	100	25	100
1. Verdünnung	1 : 25	1 : 25	1 : 25
2. Verdünnung	keine	1 : 25	1 : 25

Tabelle 6.7-4: Probenaufbereitungsschritte der Reihe C.

Tag	(5)	(3)	(4)	(6)
1	1,0	0,1	0,025	0,1
2	0,75	0,075	0,75	0,75
3	0,5	0,05	0,25	1,0
4	0,25	0,25	0,1	0,5
5	0,1	1,0	0,075	0,075
6	0,075	0,025	0,5	0,25
7	0,05	0,75	0,05	0,05
8	0,025	0,5	1,0	0,025

Tabelle 6.7-5: Zusammensetzung der zugesetzten Komponenten (in %) in den Proben der Reihe C. Die Zuordnung der einzelnen Konzentrationen erfolgte randomisiert.

6.8 Datenauswertung

Die prozentuale Wiederfindung einer Einzelmessung wurde im Rahmen der Datenauswertung als Quotient von gefundener und zugegebener Substanz multipliziert mit 100 berechnet. Die Menge der gefundenen Substanz ergibt sich aus dem Verhältnis der spezifischen Fläche einer Einzelmessung zu dem Mittelwert der spezifischen Flächen der sechs Standards (sog. Einpunktkalibrierung).

Für den Konzentrationsbereich von 120 - 7,5 % in Reihe A betrug die Konzentration der Standards 100 %. Im darauffolgenden Bereich von 7,5 - 0,025 % sowie in Reihe C wurden 1 %-Standards eingesetzt. Reihe B hatte 0,1 %-Standards.

Die Response Faktoren der Verunreinigungen wurden bestimmt aus den Mittelwerten der jeweiligen Peakflächenverhältnisse der Konzentrationen von 1,0 bis 0,5 % an drei verschiedenen Tagen. Die Werte betrugen für **(3)** 0,97, für **(4)** 1,72, für **(5)** 1,37 und für **(6)** 1,85.

Die Wiederholbarkeit wurde als relative Standardabweichung aus den sechs unabhängig aufbereiteten Proben und die Systempräzision als relative Standardabweichung aus der 6-fachen Wiederholinjektion einer dieser Proben berechnet.

Für die automatische Integration mittels der D 7000 HSM Software konnten keine Parameter gefunden werden, die allen Peak-Integrationen über diesen weiten Konzentrationsbereich gerecht geworden wären. Daher sind alle Werte manuell integriert worden. Die statistische Auswertung und die Datenverarbeitung wurden mit Microsoft Excel durchgeführt.

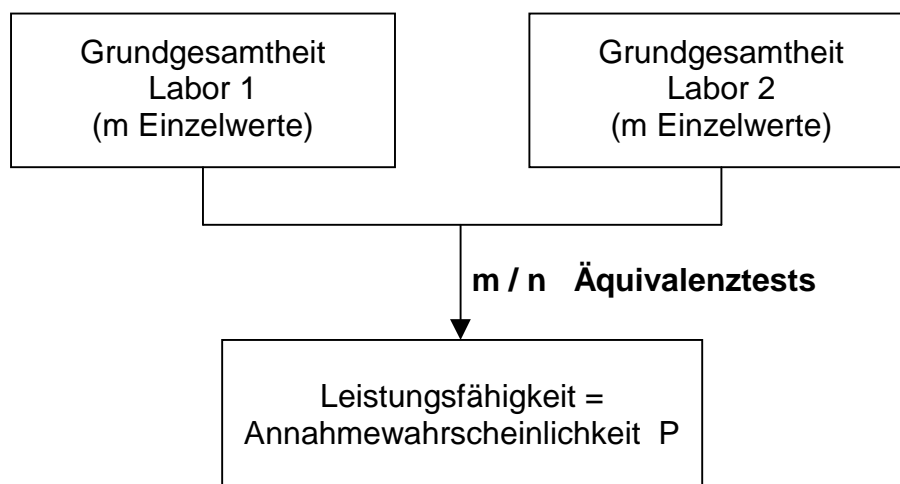
7 Simulationen und Berechnungen

7.1 Simulation des ISPE-Versuchsdesigns

7.1.1 Allgemeine Durchführung einer Simulation

Im Rahmen einer Simulation werden die zu vergleichenden Labors jeweils durch eine Grundgesamtheit repräsentiert. Eine Grundgesamtheit stellt alle diejenigen Ergebnisse dar, die bei praktisch unendlich häufiger Anwendung der Methode in dem Labor entstehen würden. Die Grundgesamtheiten der zu vergleichenden Labors besitzen untereinander den gleichen Aufbau. Allerdings unterscheiden sie sich innerhalb einer Simulation durch eine systematische Abweichung (Bias) voneinander.

An diesen Grundgesamtheiten wird anschließend der Äquivalenztest anhand der Gleichungen 2.6-5 und 2.6-6 durchgeführt. Für einen einzelnen Test wird dazu die notwendige Anzahl von Einzelwerten (Stichprobengröße n) aus beiden Grundgesamtheiten der Labors miteinbezogen, so dass sukzessive die beiden Grundgesamtheiten vollständig durchgerechnet werden. Letztlich ist innerhalb einer Simulation jeder Einzelwert der Grundgesamtheit genau einmal verwendet worden:



Die gesuchte Annahmewahrscheinlichkeit P , den Test zu akzeptieren, ergibt sich aus dem prozentualen Verhältnis an angenommen Tests zu allen durchgeführten Tests.

$$P [\%] = \frac{\text{Anzahl angenommener Tests} \cdot 100}{\text{Anzahl aller Tests}} \quad (7.1-1)$$

Sind z.B. in einer Simulation 736 von 2000 Testungen angenommen worden, so ergibt sich daraus eine Annahmewahrscheinlichkeit von:

$$P = \frac{736 \cdot 100}{2000} = 36,8\%.$$

Diese Annahmewahrscheinlichkeit ist das interessierende Ergebnis der Simulationen, die jeweils für die einzelnen Situationen durchgeführt worden sind. Ob es sich bei der Annahme eines Testes um einen Fehler 1. Art oder um eine gerechtfertigte Annahme handelt, hängt von dem Bias ab, der zwischen die beiden Grundgesamtheiten eingebaut wurde. Dieser Unterschied zwischen den Labors ist bei jeder Simulation bekannt.

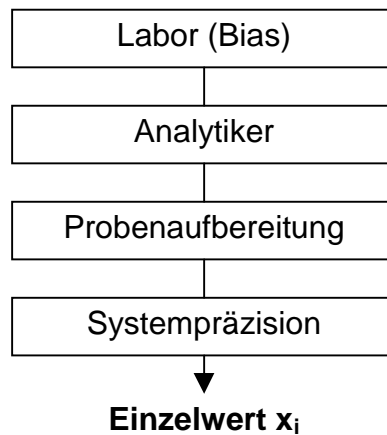
Unterscheiden sich die wahren Mittelwerte der Grundgesamtheiten in der Simulation um weniger als die Akzeptanzgrenze, dann handelt es sich bei der Annahmewahrscheinlichkeit um die Power $(1-\beta)$. Existiert zwischen den Grundgesamtheiten ein größerer Unterschied als die Akzeptanzgrenze, dann handelt es sich bei der Annahmewahrscheinlichkeit um den α -Fehler.

Auf diese Weise lässt sich die Leistungsfähigkeit der Testung in Abhängigkeit von Bias und den Varianzen bestimmen, in dem für verschiedene Konstellationen von Varianz und Bias eine Reihe von Simulationen berechnet werden.

7.1.2 Aufbau der Grundgesamtheiten

Eine Grundgesamtheit besteht jeweils aus $N = 36000$ Einzelwerten (bzw. 36000 Zeilen). Zur Erzeugung eines solchen Einzelwertes x_i werden folgende analytische Varianzkomponenten in dem Design berücksichtigt:

Analytiker $\hat{\sigma}_{an}$, Probenaufbereitung $\hat{\sigma}_{sp}$, Systempräzision $\hat{\sigma}_{sys}$ und eine systematische Abweichung zwischen den Labors (Bias).



Die obigen Varianzkomponenten werden als Faktor in den Einzelwert integriert. Ein Einzelwert x_i in einem Labor ist also das Produkt der Faktoren Analytiker $\hat{\sigma}_{an}$, Probenaufbereitung $\hat{\sigma}_{sp}$ und Systempräzision $\hat{\sigma}_{sys}$.

Die jeweiligen Faktoren ihrerseits werden durch die Excel-Funktion:

`NORMINV(ZUFALLSZAHL();1; $\hat{\sigma}$)` erzeugt.

Diese Funktion generiert normalverteilte Zufallszahlen mit einer wählbaren Standardabweichung $\hat{\sigma}$ und einem Mittelwert von 1. Sollte z.B. der Faktor $\hat{\sigma}_{sp}$ mit einer RSD von 0,60 % eingesetzt werden, so wurden die entsprechenden Zellen mit der Funktion `NORMINV(ZUFALLSZAHL();1;0,006)` belegt.

Gemäß der Hierarchie des ISPE-Designs beeinflusst der Faktor des Analytikers alle nachfolgenden Operationen, also alle Probenaufbereitungen und Messungen gleichermaßen. D.h. alle 9 Einzelwerte, die aus der Bearbeitung eines Analytikers entstehen, erhalten denselben zufälligen (Fehler)-Faktor für den Analytiker. Jeder Einzelwert unterscheidet sich dann noch in dem Faktor für die Probenaufbereitung und der Messung. Anschließend wird ein Bias (wenn notwendig) eingebaut, indem

für das Labor der Ursprungseinheit der Faktor 100 und für das Ziellabor eine Zahl mit der gewünschten Abweichung verwendet wird (z.B. 99,5 für einen Bias von 0,5). Ein Ausschnitt aus einer so generierten Grundgesamtheit ist in Abbildung 7.1-1 dargestellt.

	A	B	C	D	E	F	G	H	I	J
1	Analytiker	Probenaufb.	Systempräzision	Bias					Einzelwerte	
2	1.002	0.998	1.001	1.001	1.000	0.998	100.0	99.5	=G2*E2*	99.17
3	1.002	0.998	0.999	0.992	0.997	1.002	100.0	99.5	99.87	98.69
4	1.002	0.998	1.003	1.001	1.002	1.003	100.0	99.5	100.80	99.72
5	1.002	0.998	1.004	0.998	1.003	1.006	100.0	99.5	100.86	99.68
6	1.002	0.998	0.996	0.998	1.000	0.997	100.0	99.5	99.84	98.90
7	1.002	0.998	1.000	0.995	1.001	1.001	100.0	99.5	100.37	98.93
8	1.002	0.998	0.993	1.004	1.004	1.000	100.0	99.5	99.95	99.72
9	1.002	0.998	0.988	0.998	0.999	1.001	100.0	99.5	98.93	99.16
10	1.002	0.998	1.009	1.010	0.994	1.003	100.0	99.5	100.60	100.67
11	1.001	1.007	1.001	1.002	1.002	1.000	100.0	99.5	100.49	100.49
12	1.001	1.007	0.996	0.991	1.003	1.001	100.0	99.5	100.01	99.46
13	1.001	1.007	0.994	0.992	1.000	1.002	100.0	99.5	99.54	99.64
14	1.001	1.007	0.998	1.012	0.999	1.004	100.0	99.5	99.80	101.78
15	1.001	1.007	1.004	0.999	0.996	1.003	100.0	99.5	100.19	100.52
16	1.001	1.007	0.989	1.001	1.002	1.000	100.0	99.5	99.27	100.33
17	1.001	1.007	0.983	0.992	1.000	0.996	100.0	99.5	98.40	98.95
18	1.001	1.007	0.997	0.998	1.002	1.000	100.0	99.5	99.96	100.03
19	1.001	1.007	1.000	0.997	1.001	1.002	100.0	99.5	100.25	100.17
20	1.000	0.997	1.011	1.005	1.001	1.003	100.0	98.7	101.22	99.11

Abbildung 7.1-1: Die Einzelwerte der zu vergleichenden Labors werden in Spalte I und J generiert. Der Einzelwert in Zelle I2 setzt sich zusammen als Produkt aus den Faktoren A2, C2, E2 und G2 ($\hat{\sigma}_{an}$, $\hat{\sigma}_{sp}$, $\hat{\sigma}_{sys}$ und Bias).

Da ein Analytiker im Rahmen des ISPE-Designs aus drei Stichproben drei unabhängige Probenaufbereitungen und einer anschließenden Messung je Probenaufbereitung bearbeitet, produziert er 9 Einzelwerte. Die insgesamt 36000 Einzelwerte werden daher zu 4000 Mittelwerten der Analytiker \bar{x}_{an} zusammengefasst. Je Labor werden anfänglich 2 Analytiker eingesetzt. Das entspricht einem Datensatz mit 2000 „Methodentransfers“.

7.1.3 Berechnung der Grundgesamtheiten

Die Einzelwerte der Spalten I und J aus Abschnitt 7.1.2 werden durch den Äquivalenztest schrittweise zur Berechnung herangezogen. Die dazu benutzten Gleichungen 2.6-5 und 2.6.6 sind dabei durch die Excel-Quellcodes:

```
=(MITTELWERT(L1:L2)-MITTELWERT(K1:K2)+2)/  
WURZEL((VARIANZ(K1:K2)/2)+(VARIANZ(L1:L2)/2))  
und  
=(2-(MITTELWERT(L1:L2)-MITTELWERT(K1:K2)))/  
WURZEL((VARIANZ(K1:K2)/2)+(VARIANZ(L1:L2)/2)) gegeben.
```

Beginnend mit den Werten K1, K2, L1 und L2 (Mittelwerte der Analytiker aus den Einzelwerten der Spalten I und J (s. Abb. 7.1-1)), die als Einzelwerte in den Äquivalenztest einfließen, werden nach und nach alle Werte dieser Spalten berechnet. Die fettgedruckten Werte **+2** und **2-** in den Quellcodes sind die beiden absoluten Akzeptanzgrenzen, gegen die getestet wird und entsprechen bei einem wahren Mittelwert von 100 einer relativen Grenze von 2 %.

Null- und Alternativhypothese werden durch folgenden Quellcode wiedergegeben:

```
=WENN(UND(M2>TINV(0,1;2); N2>TINV(0,1;2));1;0).
```

Sind beide Bedingungen der UND-Verknüpfung erfüllt, wird die Alternativhypothese und damit der Transfer angenommen und eine 1 als Zelleintrag erzeugt. Die Menge aller angenommenen Tests ergibt sich dann als Summe der 1-Einträge dieser Spalte.

7.1.4 Validierung der Tests

Um sicherzustellen, dass sowohl die Formeln der angewandten Tests korrekt in die Excel-Tabellen implementiert als auch die Grundgesamtheiten korrekt erzeugt wurden, wird überprüft, ob die Tests die angegebenen Irrtumswahrscheinlichkeiten einhalten.

Für den Äquivalenztest wird dazu die Annahmewahrscheinlichkeit für den Fall berechnet, bei der die tatsächliche Abweichung zwischen den Labors gleich der Akzeptanzgrenze ist. So muss bei einer Akzeptanzgrenze von 2 % und einem

Bias von ebenfalls 2 % die Wahrscheinlichkeit den Test anzunehmen genau 5 % betragen, was in diesem Falle der gewählten Irrtumswahrscheinlichkeit α entspricht. Ist dies im Rahmen der Simulation der Fall, kann von einer korrekt erzeugten Grundgesamtheit sowie einem korrekt implementierten Test ausgegangen werden.

7.1.5 Kompromiss zwischen Leistungsfähigkeit und Präzision

Als Maß für die Leistungsfähigkeit einer Simulationen kann die Streuung der Ergebnisse zwischen den wiederholten Anwendungen der Simulation („durch die Excel-Funktion: neu berechnen“) herangezogen werden. Je höher die Datenzahlen der Simulation sind, desto länger dauern die Berechnungen und desto präziser ist das Ergebnis. Als Kompromiss zwischen Zeitbedarf und Genauigkeit wurde mit Grundgesamtheiten von 36000 Einzelwerten je Labor gerechnet. Die angegebenen Zahlenwerte für die Wahrscheinlichkeit sind demzufolge mit einer Unsicherheit (Standardabweichung) von $\pm 0,5$ verbunden.

7.1.6 Einfluss der Varianzkomponenten

Anhand der Struktur der Grundgesamtheit aus 7.1.3. sind für viele verschiedene Szenarien Datensätze generiert worden. Für den Äquivalenztest gehen mindestens je zwei \bar{x}_{an} aus beiden Labors als Einzelwerte ein. Die Mittelwerte der Analytiker \bar{x}_{an} dienen somit als Grundgesamtheiten aus denen die Äquivalenztests berechnet werden. Die Leistungsfähigkeit des Äquivalenztestes wird daher auf die Streuung der Mittelwerte der Analytiker $\hat{\sigma}_{\bar{x}_{an}}$ bezogen.

Die Streuung (als Standardabweichung) der Mittelwerte der Analytiker $\hat{\sigma}_{\bar{x}_{an}}$ berechnet sich unter Berücksichtigung des zugrundeliegenden Versuchsdesigns allgemein nach folgender Gleichung 7.1-2:

$$\hat{\sigma}_{\bar{x}_{an}} = \sqrt{\hat{\sigma}_{an}^2 + \frac{\hat{\sigma}_{sp}^2}{n_{sp}} + \frac{\hat{\sigma}_{sys}^2}{n_{sp} \cdot n_{sys}}} \quad (7.1-2)$$

Die in Abschnitt 4.2.4 betrachtete Anfangssituation mit neun unabhängigen Probenaufbereitungen ($n_{sp} = 9$), einer Messung ($n_{sys} = 1$) je Probenaufbereitung, den Werten $\hat{\sigma}_{an} = 0,30 \%$, $\hat{\sigma}_{sp} = 0,60 \%$ und $\hat{\sigma}_{sys} = 0,30 \%$ ergibt nach Gleichung 7.1-2 eine Standardabweichung von:

$$\hat{\sigma}_{\bar{x}_{an}} = \sqrt{0,30^2 + \frac{0,60^2}{9} + \frac{0,30^2}{9}} = 0,37$$

Die Grundgesamtheit der Mittelwerte der Analytiker streut in diesem Beispiel um 0,37 %. Liegt eine geringe Systempräzision vor (z.B. $\hat{\sigma}_{sys} = 1,50 \%$), so erhöht sich die Standardabweichung für die Grundgesamtheit der Mittelwerte zu:

$$\hat{\sigma}_{\bar{x}_{an}} = \sqrt{0,30^2 + \frac{0,60^2}{9} + \frac{1,50^2}{9}} = 0,62$$

In diesem Fall wäre die zu erwartende Power in Abhängigkeit von der eingesetzten Analytiker in Abbildung 4.2-2 für die $\hat{\sigma}_{\bar{x}_{an}}$ in Spalte vier abzulesen.

7.1.7 Einfluss der Verringerung der $\hat{\sigma}_{sp}$ oder der $\hat{\sigma}_{sys}$ auf die $\hat{\sigma}_{\bar{x}_{an}}$

Gemäß der oben gegebenen Fehlerstruktur besitzt die Erhöhung der Anzahl der Probenaufbereitungen oder Injektionen nur einen vergleichsweise geringen Effekt auf die Standardabweichung der Mittelwerte der Analytiker. Übt ein Analytiker bei jeder Stichprobe z.B. vier statt drei Anwendungen der Methode aus, so verringert sich die $\hat{\sigma}_{\bar{x}_{an}}$ nur auf 0,36 %.

$$\hat{\sigma}_{\bar{x}_{an}} = \sqrt{0,30^2 + \frac{0,60^2 + 0,30^2}{12}} = 0,36$$

Als Konsequenz dieses Zusammenhangs kann weder die Probenaufbereitung noch die Systempräzision die Variation der Mittelwerte der Analytiker in hohem Maße verringern, da ohnehin bereits 9 Anwendungen der Methode ausgeübt werden.

Die Erhöhung der Anzahl der Analytiker hat einen wesentlich stärkeren Effekt auf die Transfersituation. Daher wird die Abbildung 4.2-1 in Abhängigkeit von der Anzahl an Analytiker je Labor angegeben.

7.1.8 Simultane Testung dreier verschiedener Chargen anhand eines globalen Signifikanzniveaus

Die simultane Testung dreier verschiedener Chargen anhand eines globalen Signifikanzniveaus läuft wie folgt ab: Jede der drei Chargen wird einzeln getestet. Der Transfer wird nur dann angenommen, wenn alle drei Testungen akzeptiert worden sind. Um die simultane Testung mit einem globalen Signifikanzniveau von fünf Prozent durchzuführen, muss für jeden einzelnen Test die Irrtumswahrscheinlichkeit angepasst werden. Die Testungen selber laufen sonst wie in Abschnitt 2.6.4 beschrieben ab. Im Rahmen der Validierung dieser Testung anhand von nicht korrelierten Daten konnte ein globales Niveau von 5 % erhalten werden. Die Berechnung anhand des ISPE-Desings lieferte die in 4.2.5 beschriebene Verzerrung.

7.1.9 Einsatz des verbundenen t-Tests

In Abschnitt 4.2.5 ist die Interpretationsweise des ISPE-Desings in die Richtung, dass drei verschiedene Chargen einzusetzen sind, kurz erwähnt worden. Für Problemstellungen, bei denen zwei Messreihen in einer Weise voneinander abhängig sind, dass jeweils ein Messwert der einen in direkter Beziehung zu einem Messwert der zweiten Messreihe steht, ist der sogenannte verbundene t-Test einsetzbar. Die Messwerte sind durch die eingesetzten unterschiedlichen Chargen paarweise verbunden.

Die Berechnung der Prüfgröße T_t des verbundenen t-Tests in der Simulation erfolgte nach folgenden Gleichungen 7.1-3 bis 7.1-6 [25, 32]:

$$d_i = \bar{x}_{1i} - \bar{x}_{2i} \quad \text{für } i = 1, \dots, n \quad (7.1-3)$$

$$\bar{d} = (1/n) \sum_{i=1}^n d_i \quad (7.1-4) \quad \hat{\sigma}_d^2 = (1/n-1) \sum_{i=1}^n (d_i - \bar{d})^2 \quad (7.1-5)$$

$$T_t = \frac{\bar{d}}{\hat{\sigma}_d \sqrt{1/n}} \quad (7.1-6)$$

Die Prüfgröße T_t wird mit dem kritischen Wert $t_{n-1;1-\alpha/2}$ mit $n - 1$ Freiheitsgraden verglichen. Die Ergebnisse der Simulationen zeigten eine deutliche Verzerrung von α .

7.2 Berechnungen zur Testung nach USP <1010>

7.2.1 Methode

Das USP-Kapitel <1010> lässt für den Vergleich zweier Varianzen untersuchen, ob die obere Grenze des zweiseitigen 90%-Konfidenzintervalls des Quotienten zweier Varianzen R eine vorher gewählte Akzeptanzgrenze A überschreitet, oder nicht. Wobei in diesem Falle nur der Vergleich mit einer oberen Akzeptanzgrenze interessant ist, weil eine eventuell bessere Präzision des Empfängerlabors kein Problem darstellt. Die obere Grenze des Konfidenzintervalls wird nach Gleichung 2.6-13 mit der Akzeptanzgrenze verglichen.

7.2.2 Berechnung der Power

Die Berechnung der Power für eine spezielle Testsituation kann anhand folgender Excel-Funktion durchgeführt werden [46]:

$$\text{Power} = \text{FVERT}[(R/A)*\text{FINV}(\alpha, n_2-1, n_1-1), n_2-1, n_1-1].$$

R ist das tatsächliche Verhältnis zwischen den Varianzen ($\hat{\sigma}_2^2/\hat{\sigma}_1^2$) der Grundgesamtheiten für das die Power bestimmt werden soll, A ist die obere Akzeptanzgrenze, α ist die Irrtumswahrscheinlichkeit (üblicherweise = 5 %) und $n_{1,2}$ sind die Stichprobengrößen der Labors. Anhand dieser Funktion sind für die Abbildung 4.3.1 in Excel für die verschiedenen Konstellationen entsprechende Annahmewahrscheinlichkeiten berechnet worden.

7.2.3 Berechnung der Hälfte der oberen Akzeptanzgrenze

Die hintere Ebene der Abbildung 4.3.1 ist für den Fall berechnet worden, dass der Bias zwischen den beiden Grundgesamtheiten die Hälfte des zulässigen Verhältnisses der Varianzen bezogen auf die Akzeptanzgrenze A hat. Bei $A = 4$ verhalten sich die Varianzen wie 4 zu 1 bzw. die Standardabweichungen wie 2 zu 1. Das halbe zulässige Verhältnis ist dann 2 und liegt vor, wenn sich die Varianzen wie 2 zu 1 bzw. die Standardabweichungen wie $\sqrt{2}$ zu 1 (1,41 zu 1) verhalten.

Für das jeweilige A einer Spalte in Abbildung 4.3-1 berechnet sich das halbe zulässige Verhältnis allgemein nach Gl. 4.3-1. Danach ist dieses Verhältnis z.B. für den Fall $A = 9,00$ in der vierten Spalte:

$$\left[1 + \frac{(\sqrt{9} - 1)}{2}\right]^2 = \left[1 + \frac{2}{2}\right]^2 = 2^2 = 4,00$$

8 Abkürzungsverzeichnis

α	Wahrscheinlichkeit für den Fehler 1. Art
AkzG	Akzeptanzgrenze
AMG	Arzneimittelgesetz
β	Wahrscheinlichkeit für den Fehler 2. Art
$c_{1-\alpha}$	kritischer Wert
C_L	untere Grenze eines Konfidenzintervalls
C_U	obere Grenze eines Konfidenzintervalls
$F_{n1,n2,\gamma}$	kritischer Wert einer F-Verteilung
FDA	Food and Drug Administration
H_0	Nullhypothese
H_1	Alternativhypothese
HPLC	Hochleistungs-Flüssigchromatographie
LOQ	Bestimmungsgrenze
μ	Erwartungswert
N, n	Stichprobenumfang
n_{sp}	Anzahl an Probenaufbereitungen
n_{sys}	Anzahl an Messungen
ν	effektiver Freiheitsgrad
P	Wahrscheinlichkeit
θ	Parameter einer Verteilung
θ_0	Sollwert, gegen den θ getestet wird
$\theta - \varepsilon$	untere Akzeptanzgrenze
$\theta + \varepsilon$	obere Akzeptanzgrenze
R	Auflösung
RSD%	prozentuale relative Standardabweichung
σ	Standardabweichung
σ^2	Varianz
$\hat{\sigma}$	empirische Standardabweichung
$\hat{\sigma}^2$	empirische Varianz
$\hat{\sigma}_{inj}$	Injektionsfehler
$\hat{\sigma}_{int}$	Mehrtages-Standardabweichung

$\hat{\sigma}_p$	gepoolte Standardabweichung
$\hat{\sigma}_r$	Wiederhol-Standardabweichung
$\hat{\sigma}_R$	Ringversuchs-Standardabweichung
$\hat{\sigma}_{sp}$	Probenaufbereitungsfehler
$\hat{\sigma}_{sys}$	Systempräzision
$\hat{\sigma}_{\bar{x}_{an}}$	Standardabweichung der Mittelwerte der Analytiker
$\hat{\sigma}_{\bar{x}_s}$	Standardabweichung der Mittelwerte der Serien
$\hat{\sigma}_{zw}$	Variation zwischen den Serien
T	Prüfgröße
T _F	Prüfwert des F-Tests
t _{n;γ}	kritischer Wert einer t-Verteilung
T _t	Prüfwert des t-Tests
USP	United States Pharmacopeia
\bar{x}	arithmetischer Mittelwert
%AkzG	prozentualer Anteil der Präzision an der Akzeptanzgrenze

9 Literaturverzeichnis

- [1] Georg Schwedt, *Analytische Chemie*, **1995**, Georg Thieme Verlag, Stuttgart, New York, 1. Auflage
- [2] G Rücker, M Neugebauer und GG Willems, *Instrumentelle pharmazeutische Analytik*, **1992**, Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart, 2. Auflage
- [3] U Wintermeyer, *Die Wurzeln der Chromatographie*, **1989**, GIT Verlag GmbH, Darmstadt, 1. Auflage
- [4] AJP Martin und RLM Syngé, *Biochem J*, **1941**, 35, 1358-1368
- [5] M Otto, *Analytische Chemie*, **2000**, WILEY-VCH Verlag GmbH, Weinheim, 2. Auflage
- [6] J Ermer, *Journal of Pharmaceutical and Biomedical Analysis*, **2001**, 24, 755-767
- [7] Q7A: *Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients*, **2000**, verfügbar unter: www.ich.org
- [8] K Metzger, *Pharm. Ind.*, **1999**, 61, 945-952
- [9] Q2A: *Text on Validation of Analytical Procedures*, **1994**
verfügbar unter: www.ich.org
- [10] Q2B: *Validation of Analytical Procedures : Methodology*, **1996**
verfügbar unter: www.ich.org
- [11] AGJ Daas und JHMcB Miller, *Pharmeuropa*, **1999**, 11, 571-577
- [12] AGJ Daas und McB Miller JH, *Pharmeuropa*, **1998**, 10, 137-146

- [13] B Renger, *Journal of Chromatography B*, **2000**, 745, 167-176
- [14] J Ermer et al., *Journal of Pharmaceutical and Biomedical Analysis*, **2005**, 38, 653-663
- [15] AGJ Daas und JHMcB Miller, *Pharmeuropa*, **1997**, 9, 148-156
- [16] T Anglov et al., *Accerd. Qual. Assur.*, **2003**, 8, 225-230
- [17] S Küppers, B Renger und VR Meyer, *LC-GC Europe*, **2000**, February, 114-118
- [18] Y Hayashi und R Matsuda, *Analytical Chemistry*, **1994**, 66, 2874-2881
- [19] ISPE Good Practice Guide: *Technology Transfer*, **2003**
verfügbar unter: www.ispe.org
- [20] R Kringle, R Kahn-Malek, F Snikeris, P Munden, C Agut und M Bauer, *Drug information journal*, **2001**, 35, 1271-1288
- [21] J Ermer und JHMcB Miller, *Method validation in Pharmaceutical Analysis*, **2005**, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 1. Auflage
- [22] J Vial, A Jardy, P Anger, A Brun und JM Menet, *Journal of Chromatography A*, **1998**, 815, 173-182
- [23] Ph.Eur. Nachtrag 2001, **2001**, 35-40
- [24] VR Meyer, *Praxis der Hochleistungs-Flüssigchromatographie*, **1999**, Otto Salle Verlag GmbH & Co., Frankfurt am Main, 8. Auflage
- [25] J Hartung, B Elpelt und KH Klösener, *Statistik*, **1991**, R. Oldenbourg Verlag GmbH, München, 8. Auflage

- [26] H Wätzig und J Ermer, *PZ-Prisma*, **2003**, 11, 257-266
- [27] T Layloff, *American Genomic/Proteomic Technology*, **2002**, 2, 14-19
- [28] GB Limentani, MC Ringo, F Ye, ML Bergquist und EO MCSorley, *Analytical Chemistry*, **2005**, 77, 221A-226A
- [29] Spezifikationsgrenzen: Positionspapier der Deutschen Pharmazeutischen Gesellschaft zum Thema Spezifikationssetzung, *Pharmazie in unserer Zeit*, **2002**, 31, 501 - 503
- [30] JL Lozan und H Kausch, *Angewandte Statistik für Naturwissenschaftler*, **1998**, Blackwell Wissenschafts-Verlag, Berlin Wien, 2. Auflage
- [31] E Kreyszig, *Statistische Methoden und ihre Anwendungen*, **1977**, Vandenhoeck & Ruprecht, Göttingen, 6. Auflage
- [32] L Sachs, *Angewandte Statistik*, **1997**, Springer-Verlag, Berlin Heidelberg, 8. Auflage
- [33] G Clauß, FR Finze und L Partsch, *Statistik*, **2002**, Wissenschaftlicher Verlag Harri GmbH, Frankfurt am Main, 4. Auflage
- [34] R Schlittgen, *Einführung in die Statistik*, **1987**, R. Oldenbourg Verlag GmbH, München, 1. Auflage
- [35] TW O' Gorman, *Commun. Statist. - Simula.*, **1997**, 26, 1393-1411
- [36] GH Mehring, *Commun. Statist. - Theory Meth.*, **1993**, 22, 1257-1297
- [37] Y Hu und HT Karnes, *Chemometrics and Intelligent Laboratory Systems*, **1999**, 45, 257-266
- [38] M Weichert und LA Hothorn, *Commun. Statist.- Simula.*, **2002**, 31, 175-187

- [39] BL Welch, *Biometrika*, **1938**, 29, 350-362
- [40] WJ Westlake, *Journal of Pharmaceutical Sciences*, **1972**, 61, 1340-1341
- [41] DJ Schuirmann, *Biometrics*, **1981**, 37, 617.
- [42] S Wellek, *Statistische Methoden zum Nachweis von Äquivalenz*, **1994**, Fischer, Stuttgart,
- [43] B Renger, *Pharm. Ind.*, **1998**, 60, 989-995
- [44] MVA - *Method Validation in Analytics*, **2001**, Novia GmbH, Frankfurt
verfügbar unter: www.novia.de
- [45] L Sachs, *Statistische Methoden 2 Planung und Auswertung*, **1990**, Springer-Verlag, Berlin Heidelberg, 1. Auflage
- [46] USP 28 NF23, *General Chapters <1010>*, **2005**
- [47] DL Massart, BGM Vandeginste, LMC Buydens, S De Jong, PJ Lewi und J Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics Part A*, **1997**, Elsevier, Amsterdam, 1. Auflage
- [48] JW Dolan, *LC-GC Europe*, **2002**, July, 394-397
- [49] JW Dolan, *LC-GC Europe*, **2004**, June, 328-332
- [50] R Albert und W Horwitz, *Analytical Chemistry*, **1997**, 69, 789-790
- [51] W Horwitz, *Analytical Chemistry*, **1982**, 54, 67A-76A
- [52] W Horwitz und R Albert, *Journal of AOAC International*, **1996**, 79, 589-621

- [53] W Funk, V Dammann und G Donnevert, *Qualitätssicherung in der Analytischen Chemie*, **2005**, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2. Auflage
- [54] G Maldener, *Chromatographia*, **1989**, 28, 85-88
- [55] DIN-ISO 5725-2:, *Accuracy (trueness and precision) of measurements and results - Part 6*, **1994**
- [56] DIN-ISO 5725-2:, *Accuracy (trueness and precision) of measurements and results - Part 2*, **1994**
- [57] C Perrin, F Huguet, DL Massart und Y Vander Heyden, *Electrophoresis*, **2003**, 24, 2469-2480
- [58] E Nürnberg und P Surmann, *Hagers Handbuch Der Pharmazeutischen Praxis*, Band 2., **1991**, Springer, Berlin Heidelberg New York, 5. Auflage
- [59] U Schepers, J Ermer, L Preu und H Wätzig, *Journal of Chromatography B*, **2004**, 810, 111-118
- [60] H Wätzig und J Ermer, *Pharmazie in unserer Zeit*, **2003**, 32, 254-256
- [61] J Ermer, *Pharm. Ind.*, **2004**, 66, 1019-1021
- [62] RE Majors, *LC-GC Europe*, **2003**, July, 404-414
- [63] M Przybyciel und RE Majors, *LC-GC Europe*, **2002**, October, 652-657
- [64] KK Unger und E Weber, *Handbuch der HPLC*, **1995**, GIT-Verlag GmbH, Darmstadt, 1. Auflage
- [65] S Kromidas, *Practical problem solving in HPLC*, **2000**, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 1. Auflage

10 Anhang

10.1 Exemplarisches Probenaufbereitungsschema aus Reihe C

Zielkonzentrationen:

STD: (1) = 1,0 %; PROBE: (1) = 100; (5) = 0,10; (3) = 1,0; (4) = 0,075; (6) = 0,075 %

Einwaagen:

Kolbennummer		Einwaage	Substanz	Kolbengröße	Art
1)		12,5 mg	(1)	in 250mL	STD
2-8)	je	5,0 mg	(1)	in 25mL	Matrix A
9)		3,125 mg	(5)	in 25mL	0,10 %
10)		5,0 mg	(3)	in 100mL	1,0 %
11)		2,344 mg	(4)	in 25mL	0,075 %
12)		2,344 mg	(6)	in 25mL	0,075 %
13-40)		s.u.			

Verdünnungen:

- Kolben **1,9,10,11,12** mit 80%-ACN auffüllen => Ultraschallbad
- Je 1mL aus Kolben **1** in die Kolben **13-19** (25mL)
- Kolben **13-19** mit 80%-ACN auffüllen => Ultraschallbad

Kolben **13-19** = 1%-Standards

- Je 1ml aus Kolben **9** in die Kolben **20-26** (25mL)
- Je 1ml aus Kolben **11** in die Kolben **27-33** (25mL)
- Je 1ml aus Kolben **12** in die Kolben **34-40** (25mL)
- Kolben **20-40** mit 80%-ACN auffüllen => Ultraschallbad
- Je 1ml aus Kolben **20,10,27,34** in Kolben **2**
- Je 1ml aus Kolben **21,10,28,35** in Kolben **3**
- ... bis ...
- Je 1ml aus Kolben **26,10,33,40** in Kolben **8**
- Kolben **2-8** mit 80%-ACN auffüllen => Ultraschallbad

Kolben **2-8** = Proben

10.2 Tabellen der Einzelwerte aus der Messreihen A

Nachfolgend abgebildet sind alle Einzelwerte der Einwaagen, die durch manuelle Integration erhaltenen absoluten Peakflächen, der Quotient aus Peakfläche und Einwaage und die einzelnen Werte für die Wiederfindung der Proben aus Reihe A. Ab einer Konzentration von 60 % sind grundsätzlich sieben Messungen durchgeführt worden. Werte mit einer Abweichung von über 3 % bezogen auf den Mittelwert sind hellgrau markiert und wurden nicht in die Berechnungen miteinbezogen.

Reihe A - 120 % bis 100 %

Nr.	Art	Einwaage [mg]	Peakfläche	Peakfläche / Einwaage	Wiederfindung [%]
1	STD	5,079	3351860	659945	
2	STD	5,032	3259390	647733	
3	STD	5,119	3325580	649654	
4	STD	4,979	3281750	659118	
5	STD	5,190	3399564	655022	
6	STD	4,937	3211140	650423	
7	STD	-	-	-	
1	120%	6,092	4040230	663203	101,46
2	120%	6,086	4059340	666996	102,04
3	120%	6,063	4019170	662901	101,42
4	120%	6,064	3975344	655565	100,29
5	120%	6,092	4040744	663287	101,47
6	120%	6,086	4052340	665846	101,87
7	120%	-	-	-	-
Nr.	Art	Einwaage [mg]	Peakfläche	Peakfläche / Einwaage	
1	STD	5,072	3310845	652769	
2	STD	5,036	3277140	650743	
3	STD	5,036	3285880	652478	
4	STD	5,016	3251370	648200	
5	STD	5,073	3295505	649617	
6	STD	5,151	3317250	644001	
7	STD	-	-	-	
1	100%	5,036	3357500	666700	102,63
2	100%	5,033	Inj.-Fehler	-	-
3	100%	5,127	3397860	662738	102,02
4	100%	5,047	3319880	657793	101,26
5	100%	5,101	3359810	658657	101,39
6	100%	5,060	3362010	664429	102,28
7	100%	-	-	-	-

Reihe A - 80 % bis 40 %

Nr.	Art	Einwaage [mg]	Peakfläche	Peakfläche / Einwaage	Wiederfindung [%]
1	STD	5,052	3401610	673319	
2	STD	5,070	3286970	648318	
3	STD	5,087	3327190	654057	
4	STD	5,005	3260560	651461	
5	STD	5,046	3297120	653413	
6	STD	5,147	3336265	648196	
7	STD	-	-	-	
1	80%	4,025	2633755	654349	100,50
2	80%	4,008	2629870	656155	100,78
3	80%	4,084	2697880	660597	101,46
4	80%	3,939	2570780	652648	100,24
5	80%	4,109	2699470	656965	100,90
6	80%	3,929	2619575	666728	102,40
7	80%	-	-	-	-
Nr.	Art	Einwaage [mg]	Peakfläche	Peakfläche / Einwaage	Wiederfindung [%]
1	STD	5,041	3378955	670295	
2	STD	4,926	3196000	648802	
3	STD	5,148	3369610	654547	
4	STD	5,167	3417790	661465	
5	STD	4,960	3239150	653054	
6	STD	4,896	3188510	651248	
7	STD	5,043	3259935	646428	
1	60%	3,100	2175563	701795	-
2	60%	3,054	2023910	662708	101,55
3	60%	3,067	2043430	666263	102,10
4	60%	3,070	2133980	695107	-
5	60%	3,116	2091250	671133	102,84
6	60%	3,057	2024530	662260	101,48
7	60%	3,066	2035500	663894	101,73
Nr.	Art	Einwaage [mg]	Peakfläche	Peakfläche / Einwaage	Wiederfindung [%]
1	STD	4,937	3233170	654886	
2	STD	4,966	3317680	668079	
3	STD	5,089	3304640	649369	
4	STD	4,941	3261775	660145	
5	STD	5,096	3358070	658962	
6	STD	5,038	3293765	653784	
7	STD	5,172	3289660	636052	
1	40%	2,047	1349995	659499	100,30
2	40%	2,018	1329750	658944	100,21
3	40%	1,993	1322730	663688	100,94
4	40%	2,082	1377510	661628	100,62
5	40%	1,921	1267370	659745	100,34
6	40%	2,038	1342140	658557	100,16
7	40%	1,929	1286215	666778	101,41

Reihe A - 20 % bis 7,5 %

Nr.	Art	Einwaage [mg]	Peakfläche	Peakfläche / Einwaage	Wiederfindung [%]
1	STD	4,994	3434810	687787	
2	STD	4,956	3204120	646513	
3	STD	5,125	3349950	653649	
4	STD	5,006	3255620	650344	
5	STD	4,935	3235590	655641	
6	STD	4,950	3237600	654061	
7	STD	5,108	3318645	649696	
1	20%	1,034	687730	665116	102,07
2	20%	1,032	679950	658866	101,11
3	20%	1,024	690670	674482	103,50
4	20%	0,989	663670	671052	102,98
5	20%	0,933	614310	658424	101,04
6	20%	1,037	678725	654508	100,44
7	20%	0,970	641475	661314	101,48
Nr.	Art	Einwaage [mg]	Peakfläche	Peakfläche / Einwaage	Wiederfindung [%]
1	STD	5,056	3335860	659782	
2	STD	5,061	3312950	654604	
3	STD	5,074	3314890	653309	
4	STD	4,993	3281705	657261	
5	STD	4,940	3228030	653447	
6	STD	4,962	3258260	656642	
7	STD	4,980	3284245	659487	
1	10%	0,555	371680	669694	102,03
2	10%	0,487	325290	667947	101,76
3	10%	0,547	364740	666801	101,59
4	10%	0,498	430060	863574	-
5	10%	0,563	374540	665258	101,36
6	10%	0,442	296560	670950	102,22
7	10%	0,523	352250	673518	102,61
Nr.	Art	Einwaage [mg]	Peakfläche	Peakfläche / Einwaage	Wiederfindung [%]
1	STD	5,039	3266740	648291	
2	STD	5,062	3275900	647155	
3	STD	5,000	3218970	643794	
4	STD	4,989	3223090	646039	
5	STD	4,946	3166140	640142	
6	STD	4,968	3217620	647669	
7	STD	4,962	3205340	645977	
1	7,5%	0,407	272360	669189	103,66
2	7,5%	0,345	233150	675797	104,68
3	7,5%	0,334	217570	651407	100,90
4	7,5%	0,384	258940	674323	104,45
5	7,5%	0,379	249350	657916	101,91
6	7,5%	0,362	233420	644807	99,88
7	7,5%	0,347	225120	648761	100,49

Reihe A - 7,5 % bis 2,5 %

Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wiederfindung [%]
1	STD	0,0499	33420	670223	
2	STD	0,0499	32105	643851	
3	STD	0,0499	32740	656586	
4	STD	0,0499	32660	654982	
5	STD	0,0499	32260	646960	
6	STD	0,0499	32210	645957	
7	STD	0,0499	31990	641545	
1	7,5%	0,3748	255180	680916	104,52
2	7,5%	0,3748	Inj.-Fehler	-	-
3	7,5%	0,3748	255315	681276	104,58
4	7,5%	0,3748	254675	679568	104,32
5	7,5%	0,3748	251480	671043	103,01
6	7,5%	0,3748	253500	676433	103,84
7	7,5%	-	-	-	-
Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wiederfindung [%]
1	STD	0,0502	32725	651996	
2	STD	0,0502	Inj.-Fehler	-	
3	STD	0,0502	32180	641138	
4	STD	0,0502	32450	646517	
5	STD	0,0502	31940	636356	
6	STD	0,0502	32220	641935	
7	STD	0,0502	32815	653789	
1	5,0%	0,2411	163585	678438	105,14
2	5,0%	0,2411	160380	665146	103,08
3	5,0%	0,2411	165175	685032	106,16
4	5,0%	0,2411	164865	683747	105,96
5	5,0%	0,2411	163320	677339	104,97
6	5,0%	0,2411	160715	666535	103,29
7	5,0%	-	-	-	-
Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wiederfindung [%]
1	STD	0,0500	31500	630202	
2	STD	0,0500	32560	651408	
3	STD	0,0500	32450	649208	
4	STD	0,0500	31620	632602	
5	STD	0,0500	32835	656910	
6	STD	0,0500	32380	647807	
7	STD	-	-	-	
1	2,5%	0,1249	83545	668745	103,73
2	2,5%	0,1249	83985	672267	104,28
3	2,5%	0,1249	85000	680392	105,54
4	2,5%	0,1249	83995	672347	104,29
5	2,5%	0,1249	84630	677430	105,08
6	2,5%	0,1249	84400	675589	104,79
7	2,5%	-	-	-	-

Reihe A - 1,0 % bis 0,50 %

Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wiederfindung [%]
1	STD	0,0498	32700	656047	
2	STD	0,0498	32510	652235	
3	STD	0,0498	32855	659157	
4	STD	0,0498	32250	647019	
5	STD	0,0498	31510	632172	
6	STD	0,0498	32050	643006	
7	STD	0,0498	32790	657852	
1	1,0%	0,0496	32405	653748	100,63
2	1,0%	0,0496	32620	658086	101,30
3	1,0%	0,0496	33465	675133	103,92
4	1,0%	0,0496	32435	654354	100,73
5	1,0%	0,0496	32890	663533	102,14
6	1,0%	0,0496	32310	651832	100,34
7	1,0%	-	-	-	-
Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wiederfindung [%]
1	STD	0,0496	31600	637097	
2	STD	0,0496	32220	649597	
3	STD	0,0496	32440	654032	
4	STD	0,0496	32290	651008	
5	STD	0,0496	32705	659375	
6	STD	0,0496	32300	651210	
7	STD	0,0496	31890	642944	
1	0,75%	0,0374	25010	668574	102,96
2	0,75%	0,0374	24160	645851	99,47
3	0,75%	0,0374	25350	677663	104,36
4	0,75%	0,0374	24485	654539	100,80
5	0,75%	0,0374	24540	656009	101,03
6	0,75%	0,0374	25760	688623	106,05
7	0,75%	0,0374	24640	658683	101,44
Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wiederfindung [%]
1	STD	0,0504	32495	645254	
2	STD	0,0504	32350	642375	
3	STD	0,0504	32090	637212	
4	STD	0,0504	32360	642573	
5	STD	0,0504	32515	645651	
6	STD	0,0504	32455	644460	
7	STD	0,0504	31940	634234	
1	0,50%	0,0251	16220	646215	100,71
2	0,50%	0,0251	16095	641235	99,93
3	0,50%	0,0251	16490	656972	102,38
4	0,50%	0,0251	16320	650199	101,33
5	0,50%	0,0251	17060	679681	105,92
6	0,50%	0,0251	16570	660159	102,88
7	0,50%	0,0251	16480	656574	102,32

Reihe A - 0,25 % bis 0,075 %

Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wieder- findung [%]
1	STD	0,0501	32390	646455	
2	STD	0,0501	32850	655636	
3	STD	0,0501	32440	647453	
4	STD	0,0501	33600	670605	
5	STD	0,0501	32720	653042	
6	STD	0,0501	32140	641466	
7	STD	0,0501	31790	634480	
1	0,25%	0,0124	9140	734254	112,98
2	0,25%	0,0124	8200	658740	101,36
3	0,25%	0,0124	9150	735058	113,11
4	0,25%	0,0124	8080	649100	99,88
5	0,25%	0,0124	8540	686054	105,57
6	0,25%	0,0124	8695	698506	107,48
7	0,25%	0,0124	8440	678021	104,33
Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wieder- findung [%]
1	STD	0,0497	31430	632445	
2	STD	0,0497	32230	648543	
3	STD	0,0497	31140	626610	
4	STD	0,0497	31740	638683	
5	STD	0,0497	32140	646732	
6	STD	0,0497	32345	650857	
7	STD	0,0497	32185	647638	
1	0,10%	0,0050	3490	697833	108,76
2	0,10%	0,0050	3290	657842	102,52
3	0,10%	0,0050	3280	655843	102,21
4	0,10%	0,0050	3100	619851	96,60
5	0,10%	0,0050	3335	666840	103,93
6	0,10%	0,0050	3370	673838	105,02
7	0,10%	0,0050	3215	642846	100,19
Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wieder- findung [%]
1	STD	0,0499	31710	635726	
2	STD	0,0499	32555	652666	
3	STD	0,0499	32230	646151	
4	STD	0,0499	32750	656576	
5	STD	0,0499	31525	632017	
6	STD	0,0499	32735	656275	
7	STD	0,0499	32440	650361	
1	0,075%	0,0037	2500	667307	103,12
2	0,075%	0,0037	2630	702007	108,48
3	0,075%	0,0037	2435	649957	100,44
4	0,075%	0,0037	2545	679319	104,98
5	0,075%	0,0037	2355	628603	97,14
6	0,075%	0,0037	2340	624600	96,52
7	0,075%	0,0037	2390	637946	98,58

Reihe A - 0,050 % bis 0,025 %

Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wieder- findung [%]
1	STD	0,0500	32175	643706	
2	STD	0,0500	32785	655910	
3	STD	0,0500	32015	640505	
4	STD	0,0500	31930	638804	
5	STD	0,0500	32400	648207	
6	STD	0,0500	32380	647807	
7	STD	0,0500	32080	641805	
1	0,050%	0,0025	1570	632962	98,10
2	0,050%	0,0025	1445	582567	90,29
3	0,050%	0,0025	1265	509998	79,04
4	0,050%	0,0025	1280	516046	79,98
5	0,050%	0,0025	1850	745847	115,59
6	0,050%	0,0025	1430	576520	89,35
7	0,050%	0,0025	1610	649089	100,60
Nr.	Art	Einwaage [mg]	Fläche	Fläche / Einwaage	Wieder- findung [%]
1	STD	0,0498	33030	662614	
2	STD	0,0498	32365	649274	
3	STD	0,0498	32135	644660	
4	STD	0,0498	32130	644559	
5	STD	0,0498	32980	661611	
6	STD	0,0498	32560	653186	
7	STD	0,0498	32060	643155	
1	0,025%	0,0012	760	613893	94,26
2	0,025%	0,0012	835	674475	103,56
3	0,025%	0,0012	770	621971	95,50
4	0,025%	0,0012	780	630048	96,74
5	0,025%	0,0012	820	662359	101,70
6	0,025%	0,0012	500	403877	62,01
7	0,025%	0,0012	735	593700	91,16

11 Lebenslauf

Persönliche Daten

Name	Udo Schepers
Geburtsdatum	02.05.1974
Geburtsort	Papenburg
Familienstand	verheiratet
Staatsangehörigkeit	deutsch

Schulbildung

08.1981 – 07.1985	Grundschule Papenburg
08.1985 – 07.1987	Orientierungsstufe Papenburg
08.1987 – 07.1994	Gymnasium Papenburg (Abitur)

Grundwehrdienst	ABC-Abwehrbataillon Emden
------------------------	---------------------------

Studium

10.1995 – 03.1996	Lehramt Mathematik/Chemie Universität Münster
04.1996 – 05.2000	Pharmazie TU-Braunschweig

Beruflicher Werdegang

07.2000 – 06.2001	Praktische Ausbildung „Apotheke am Bankplatz“ in Braunschweig
29.06.2001	Approbation als Apotheker
seit 07.2001	Wissenschaftlicher Mitarbeiter am Institut für Pharmazeutische Chemie der Technischen Universität Braunschweig